1982

Statistical computing support for Lp estimation in augmented linear models under linear inequality restrictions

Char-Lung (Charles) Lin
Iowa State University

Lin, Char-Lung (Charles)

STATISTICAL COMPUTING SUPPORT FOR L(,P) ESTIMATION IN AUGMENTED LINEAR MODELS UNDER LINEAR INEQUALITY RESTRICTIONS

Iowa State University PH.D. 1982

Statistical computing support for ℓ_p estimation in augmented linear models under linear inequality restrictions

by

Char-Lung (Charles) Lin

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1982

TABLE OF CONTENTS

1. INTRODUCTION

The linear least squares estimation problem is to minimize the sum of squared residuals while fitting the data with a linear model. Let $\underline{y} = X\underline{\beta} + \underline{e}$ be the given linear model, where $\underline{y}$ is the data vector, X is the design matrix, $\underline{\beta}$ is the parameter vector to be estimated, and $\underline{e}$ is the error vector. Least squares estimates are known as best linear unbiased estimates when $E(\underline{e}) = \underline{0}$ and $\mathrm{Var}(\underline{e}) = \sigma^2 I$, where $\sigma^2 > 0$ is a constant and I is an identity matrix. Also, least squares estimates are easy to compute. Hence, least squares estimation has been applied extensively in a variety of disciplines of science.

In some situations minimizing the ℓ_p norm of the error vector in the linear model with $p \geq 1$ and $p \neq 2$ can provide estimates that are better in some ways than the least squares estimates. Also, the amount of computation needed to obtain least ℓ_p norm estimates (often called simply ℓ_p estimates) has been greatly reduced by employing some of the new techniques developed in recent years. Hence, least ℓ_p norm estimation (often called simply ℓ_p estimation) is playing a more important role than ever before in the linear model fitting. We now give a literature review of ℓ_p estimation in the linear model with emphasis on the robustness of the ℓ_p estimates and other justification for its application. Additional literature specific to computing ℓ_p estimates will be reviewed in subsequent chapters.

Huber (1964) gave some results indicating that ℓ_p estimation

for $1 < p < 2$ is more robust than the least squares estimation when the errors in the linear model follow a nonnormal or contaminated normal distribution. Forsythe (1972) demonstrated the robustness of ℓ_p estimation in the linear model, $1 < p < 2$, based on the mean square error criterion. He compared ℓ_p estimation ($p = 1.25, 1.5, 1.75$) with the least squares estimation for the linear model in a Monte Carlo study and found that the result obtained using ℓ_p estimation is substantially better than those which used the least squares estimation when the error distribution is contaminated to produce long-tailed or skewed residuals, and is not very bad when the error distribution is truly normal. He suggested that the value $p = 1.5$ seems to be a good "compromise".

Rice and White (1964) discussed ℓ_1 estimation in the linear model and pointed out the feature of its resistence to outliers in the data and to heavy-tailed error distribution. It is also known that the ℓ_1 estimates are the maximum likelihood estimates when the errors in the linear model follow a double exponential distribution which is long-tailed and has kurtosis of 6. Harter (1977) suggested an adaptive procedure using ℓ_1 estimation if kurtosis of the error distribution in the linear model is larger than 3.8. Since there is a high probability of having kurtosis larger than 3.8 when the underlying distribution has long tails or outliers are presented because of contamination, the result of using ℓ_1 estimation in Harter's adaptive procedure will not be distorted by outliers or values far out in the tails.

Money et al. (1982) conducted a simulation study of the ℓ_p estimation ($p = 1.0, 1.25, 1.5, 1.75, 2, \infty$) in the linear model. They generated the sample data using symmetric error distribution of various kurtosis, computed the respective ℓ_p estimates, and made some comparisons based on the empirical generalized variance of the estimates. They proposed a formula for the choice of p in the ℓ_p estimation problem such that

$$p = \frac{9}{k^2} + 1 ,$$

where k is the actual or estimated population kurtosis. They concluded that their method is generally superior to either the least squares or Harter's adaptive procedure. They also indicated that there is an evidence of unbiasedness of the ℓ_p estimates when the error distribution in the linear model is symmetric, based on the fact that all sample means of the estimates are "close" to the true parameter values.

ℓ_p estimation in the linear model does not provide a unique ℓ_p estimate in general. In case there is a unique ℓ_p estimate, the unbiasedness of the ℓ_p estimate is assured when the error distribution in the linear model is symmetric. Harvey (1978) provided a simple proof for $p > 1$. Sielken and Hartley (1973) also indicated this fact for $p = 1$ or ∞ . Money et al. (1982) imposed conditions on the sample data generated for their simulation study to assure the uniqueness of ℓ_p estimate. Hence, they found the ℓ_p estimates are unbiased in their simulation study. In case the uniqueness of ℓ_p

estimate is not guaranteed, we can find a ℓ_p estimate which is unbiased when the error distribution in the linear model is symmetric. Sielken and Hartley (1973) proposed a computational algorithm using an unbiased antisymmetrical ℓ_p estimate, e.g., the least squares estimate, to construct an unbiased ℓ_p estimate, $p = 1$ or ∞. Sposito (1982) extended Sielken and Hartley's computational scheme to ℓ_p estimation, $p > 1$, in the linear model with symmetrical error distribution.

We now formally introduce the ℓ_p estimation problem in the notation used in subsequent chapters. Let $\|\underline{v}\|_p$ denote the ℓ_p norm of the vector $\underline{v}$ in $\mathbb{R}^n$, i.e., $\|\underline{v}\|_p = (\sum_{i=1}^{n} |v_i|^p)^{\frac{1}{p}}$, where $\underline{v} = (v_1, v_2, \ldots, v_n)^T$ and $p \geq 1$. Let us consider the linear model discussed earlier, $\underline{y} = X\underline{\beta} + \underline{e}$, where $\underline{y} \in \mathbb{R}^n$, $\underline{\beta} \in \mathbb{R}^m$, $\underline{e} \in \mathbb{R}^n$, and X is of dimension $n \times m$ with $n \geq m$. Let rank $(X) = t \leq m$. We need to find a $\underline{\beta}^* \in \mathbb{R}^m$ such that $\|\underline{y} - X\underline{\beta}^*\|_p \leq \|\underline{y} - X\underline{\beta}\|_p$ for all $\underline{\beta} \in \mathbb{R}^m$, where $p \geq 1$. The $\underline{\beta}^*$ found is referred to as an ℓ_p estimate. Note that, $\|\underline{y} - X\underline{\beta}^*\|_p \leq \|\underline{y} - X\underline{\beta}\|_p$ for all $\underline{\beta} \in \mathbb{R}^m$, if and only if $\|\underline{y} - X\underline{\beta}^*\|_p^p \leq \|\underline{y} - X\underline{\beta}\|_p^p$ for all $\underline{\beta} \in \mathbb{R}^m$. Hence, we can let

$$F(\underline{\beta}) = \|\underline{y} - X\underline{\beta}\|_p^p$$

$$= \sum_{i=1}^{n} |\underline{y} - \underline{x}_i^T\underline{\beta}|^p ,$$

where $\underline{x}_i^T$ is the i^{th} row in X, be the objective function for the ℓ_p estimation problem. Note that, in the linear model, the ℓ_p estimates are the least absolute estimates when $p = 1$ and are the least squares estimates when $p = 2$.

Adding the linear inequality restrictions, $A\underline{\beta} \geq \underline{b}$, where A is a matrix of dimension r x m, $r \leq m$, and $\underline{b} \in \mathbb{R}^r$, to the ℓ_p estimation problem, we get a constrained ℓ_p estimation problem. In other words, we need to find a $\underline{\beta}^* \in \mathbb{R}^m$ such that $A\underline{\beta}^* \geq \underline{b}$ and $F(\underline{\beta}^*) \leq F(\underline{\beta})$ for all $\underline{\beta} \in \mathbb{R}^m$ such that $A\underline{\beta} \geq \underline{b}$. We would assume there exists a $\underline{\beta}^* \in \mathbb{R}^m$ such that $A\underline{\beta}^* \geq \underline{b}$ to assure the feasibility of the solution.

2. SOME BASIC PROPERTIES OF THE ℓ_p ESTIMATION PROBLEM

By the nature of the linear model fitting, when the design matrix is not full-rank, we can construct another linear model with full-rank design matrix. Then, ℓ_p estimates of the parameters in the original linear model can be retrieved from an ℓ_p estimate of the parameters in the linear model with full-rank design matrix. Also, the set of ℓ_p estimates in the original linear model is a hypersubspace in $\mathbb{R}^m$. For the simplicity of the discussion on ℓ_p estimation, and without substantial loss of generality, we will assume the design matrix X in Chapter 1 is full-rank unless otherwise specified. Details of the relationship between full and nonfull-rank models are given in Section 2.1. Convexity of the objective function is a useful property. We will prove this result in Section 2.2. Finally, the existence and uniqueness of ℓ_p estimates are discussed in Section 2.3.

2.1 Rank of the Design Matrix

We can assume that the design matrix X in Chapter 1 is full-rank without loss of generality. We are reasoning as follows. Suppose that rank $(X) = t < m$. Let X be decomposed as

$$X = Q \begin{pmatrix} R_1 & 0 \\ 0 & 0 \end{pmatrix} U^T ,$$

where Q, U, and R_1 are matrices of dimension $n \times n$, $m \times m$, and $t \times t$ respectively, such that $Q^TQ = QQ^T = I$, $U^TU = UU^T = I$, and R_1 is nonsingular and is in a lower triangular form. Then,

$$X\underline{\beta} = Q\begin{pmatrix} R_1 & 0 \\ 0 & 0 \end{pmatrix} U^T\underline{\beta}$$

$$= Q\begin{pmatrix} R_1 & 0 \\ 0 & 0 \end{pmatrix} \underline{a}$$

where $\underline{a} = U^T\underline{\beta}$. Let $\underline{a} = \begin{pmatrix} \underline{a}_1 \\ \underline{a}_2 \end{pmatrix}$, where $\underline{a}_1 \in \mathbb{R}^t$ and $\underline{a}_2 \in \mathbb{R}^{m-t}$. Let $Q = (Q_1 Q_2)$, where Q_1 and Q_2 are matrices of dimension n x t and n x (n-t) respectively. Now,

$$Q\begin{pmatrix} R_1 & 0 \\ 0 & 0 \end{pmatrix} \underline{a} = (Q_1 R_1 \quad 0) \begin{pmatrix} \underline{a}_1 \\ \underline{a}_2 \end{pmatrix}$$

$$= Q_1 R_1 \underline{a}_1$$

$$= S_1 \underline{a}_1$$

where $S_1 = Q_1 R_1$, a matrix of dimension n x t .

$$I = Q^T Q$$

$$= \begin{pmatrix} Q_1^T \\ Q_2^T \end{pmatrix} (Q_1 Q_2)$$

$$= \begin{pmatrix} Q_1^T Q_1 & Q_1^T Q_2 \\ Q_2^T Q_1 & Q_2^T Q_2 \end{pmatrix}$$

implies $Q_1^T Q_1 = I$. Now, rank (S_1) = rank $(S_1^T S_1)$ = rank $(R_1^T Q_1^T Q_1 R_1)$ = rank $(R_1^T R_1)$ = rank (R_1) = t . Hence, S_1 is full-rank. Therefore, we can work with a full-rank linear model

$$\underline{y} = S_1 \underline{a}_1 + \underline{e}$$

Note that ℓ_p estimates in the original linear model can be retrieved by

$$\underline{\beta}^* = U\begin{pmatrix}\underline{a}_1^* \\ \underline{a}_2\end{pmatrix}$$

where $\underline{a}_1^*$ is an ℓ_p estimate in the full-rank linear model and $\underline{a}_2$ is an arbitrary vector in $\mathbb{R}^{m-t}$. Further, let $U = (U_1 U_2)$, where U_1 and U_2 are matrices of dimension $m \times t$ and $m \times (m-t)$ respectively, then

$$\begin{pmatrix}U_1^T \\ U_2^T\end{pmatrix} \underline{\beta}^* = U^T U \begin{pmatrix}\underline{a}_1^* \\ \underline{a}_2\end{pmatrix}$$

$$= \begin{pmatrix}\underline{a}_1^* \\ \underline{a}_2\end{pmatrix}$$

i.e.,

$$U_1^T \underline{\beta}^* = \underline{a}_1^*$$

Thus, the set of ℓ_p estimates in the original linear model is the hypersubspace defined by

$\{\underline{\beta}^* \in \mathbb{R}^m | U_1^T \underline{\beta}^* = \underline{a}_1^*$, where $\underline{a}_1^*$ is an ℓ_p estimate in the full-rank linear model $\underline{y} = S_1 \underline{a}_1 + \underline{e}\}$.

2.2 Convexity of the Objective Function

The fact that the objective function of the ℓ_p estimation problem is convex simplifies the computing problem. In fact, the objective function is strictly convex when $p > 1$ and X is full-rank. We will discuss the case of $p = 1$ first and prove convexity of $F(\underline{\beta})$.

Theorem 2.2.1. When $p = 1$, $F(\underline{\beta}) = \sum_{i=1}^{n} |y_i - \underline{x}_i^T \underline{\beta}|$ is convex.

Proof: Let $\underline{\beta}_1, \underline{\beta}_2 \in \mathbb{R}^m$ such that $\underline{\beta}_1 \neq \underline{\beta}_2$. Let $0 < \alpha < 1$.

$$F(\alpha\underline{\beta}_1 + (1-\alpha)\underline{\beta}_2) = \sum_{i=1}^{n} |y_i - \underline{x}_i^T(\alpha\underline{\beta}_1 + (1-\alpha)\underline{\beta}_2)|$$

$$= \sum_{i=1}^{n} |\alpha(y_i - \underline{x}_i^t\underline{\beta}_1) + (1-\alpha)(y_i - \underline{x}_i^T\underline{\beta}_2)| .$$

$$\leq \sum_{i=1}^{n} (|\alpha(y_i - \underline{x}_i^T \underline{\beta}_1)| + |(1-\alpha)(y_i - \underline{x}_i^T\underline{\beta}_2)|)$$

$$= \alpha\sum_{i=1}^{n} |y_i - \underline{x}_i^T\underline{\beta}_1| + (1-\alpha) \sum_{i=1}^{n} |y_i - \underline{x}_i^T\underline{\beta}_2|$$

$$= \alpha F(\underline{\beta}_1) + (1-\alpha)\, F(\underline{\beta}_2)$$

Thus, by definition, $F(\underline{\beta})$ is convex. □

In the case of $p > 1$, we will first introduce some results which are needed in the proof of the main theorem.

Lemma 2.2.1. Let $g(x) = x^p$, where $x \geq 0$ and $p > 1$, then $g(x)$ is a strictly increasing and strictly convex function of x.

Proof: Since $\frac{\partial}{\partial x} g(x) = p\, x^{p-1} > 0$ when $x > 0$, $g(x)$ is strictly increasing when $x > 0$. Also, $g(0) = 0$ and $g(x) > 0$ when $x > 0$, hence, $g(x)$ is a strictly increasing function of x, where $x \geq 0$. Note that $\frac{\partial}{\partial x} g(x)$ exists and is continuous on $[0,\infty)$. $\frac{\partial^2}{\partial x^2} g(x) = p(p-1)x^{p-2}$, hence, $\frac{\partial^2}{\partial x^2} g(x)$ exists on $(0,\infty)$. By Taylor's theorem,

$$g(a) = g(b) + \frac{\partial}{\partial x} g(b)(a-b) + \frac{1}{2}\frac{\partial^2}{\partial x^2} g(c)(a-b)^2$$

where $a \geq 0$, $b \geq 0$, $a \neq b$, and c lies between a and b . Since

$$\frac{1}{2}\frac{\partial^2}{\partial x^2} g(c)(a-b)^2 = \frac{1}{2} p(p-1)c^{p-2}(a-b)^2$$

$$> 0$$

we have

$$g(a) - g(b) > \frac{\partial}{\partial x} g(b)(a-b) .$$

Now let $x_1 \geq 0$ and $x_2 \geq 0$ such that $x_1 \neq x_2$. Let $0 < \alpha < 1$. Let $a = x_1$ and $b = \alpha x_1 + (1-\alpha)x_2$. We have

$$g(x_1) - g(\alpha x_1 + (1-\alpha)x_2) > \frac{\partial}{\partial x} g(\alpha x_1 + (1-\alpha)x_2)(1-\alpha)(x_1 - x_2) .$$

Let $a = x_2$ and $b = \alpha x_1 + (1-\alpha)x_2$. We have

$$g(x_2) - g(\alpha x_1 + (1-\alpha)x_2) > \frac{\partial}{\partial x} g(\alpha x_1 + (1-\alpha)x_2)(-\alpha)(x_1 - x_2) .$$

Multiplying the first inequality by α and the second inequality by $1-\alpha$, then adding them up, we get

$$g(\alpha x_1 + (1-\alpha)x_2) < \alpha g(x_1) + (1-\alpha)g(x_2) .$$

Thus, by definition, g is strictly convex. □

Corollary 2.2.1. Let $h(x) = |x|^p$, where $x \in \mathbb{R}$ and $p > 1$, then $h(x)$ is a strictly convex function of x .

Proof: Let $x_1 \in \mathbb{R}$ and $x_2 \in \mathbb{R}$ such that $x_1 \neq x_2$. Let $0 < \alpha < 1$. Note that $h(x) = g(|x|)$, where g is as in Lemma 2.2.1. Since

$|\alpha x_1 + (1-\alpha)x_2| \leq \alpha|x_1| + (1-\alpha)|x_2|$ and, by Lemma 2.2.1, g is increasing, we have

$$h(\alpha x_1 + (1-\alpha)x_2) = g(|\alpha x_1 + (1-\alpha)x_2|)$$
$$\leq g(\alpha|x_1| + (1-\alpha)|x_2|) .$$

In case of $|x_1| \neq |x_2|$, by Lemma 2.2.1, g is strictly convex, we have

$$g(\alpha|x_1| + (1-\alpha)|x_2|) < \alpha g(|x_1|) + (1-\alpha)g(|x_2|)$$
$$= \alpha h(x_1) + (1-\alpha)h(x_2) .$$

Hence, $h(\alpha x_1 + (1-\alpha)x_2) < \alpha\, h(x_1) + (1-\alpha)h(x_2)$.

In case of $x_1 = -x_2$,

$$h(\alpha x_1 + (1-\alpha)x_2) = h((2\alpha-1)x_1)$$
$$= g(|(2\alpha-1)x_1|)$$

and $$\alpha h(x_1) + (1-\alpha)h(x_2) = \alpha g(|x_1|) + (1-\alpha)g(|x_1|)$$
$$= g(|x_1|) .$$

Since $|2\alpha-1| < 1$ and, by Lemma 2.2.1, g is strictly increasing,

$$g(|(2\alpha-1)x_1|) < g(|x_1|)$$

i.e., $$h(\alpha x_1 + (1-\alpha)x_2) < \alpha h(x_1) + (1-\alpha)h(x_2) .$$

Thus, by definition, h is strictly convex. □

Corollary 2.2.2. Let $\underline{e} \in \mathbb{R}^n$ and $k(\underline{e}) = \sum_{i=1}^{n} |e_i|^p$, $p > 1$. Then $k(\underline{e})$ is a strictly convex function of $\underline{e}$.

Proof: Let $\underline{e}^{(1)} \in \mathbb{R}^n$ and $\underline{e}^{(2)} \in \mathbb{R}^n$ such that $\underline{e}^{(1)} \neq \underline{e}^{(2)}$. Let $I = \{1 \leq i \leq n | e_i^{(1)} = e_i^{(2)}\}$ and $J = \{1, 2, \ldots, n\} - I$. Let $0 < \alpha < 1$. Then,

$$k(\alpha\underline{e}^{(1)} + (1-\alpha)\underline{e}^{(2)}) = \sum_{i\in I} |e_i^{(1)}|^p + \sum_{i\in J} |\alpha e_i^{(1)} + (1-\alpha)e_i^{(2)}|^p .$$

By Corollary 2.2.1,

$$|\alpha e_i^{(1)} + (1-\alpha)e_i^{(2)}|^p < \alpha |e_i^{(1)}|^p + (1-\alpha)|e_i^{(2)}|^p , \quad i \in J .$$

Thus, $k(\alpha\underline{e}^{(1)} + (1-\alpha)\underline{e}^{(2)}) < \sum_{i\in I} |e_i^{(1)}|^p + \sum_{i\in J} (\alpha|e_i^{(1)}|^p + (1-\alpha)\, e_i^{(2)}|^p)$

$$= \alpha \sum_{i=1}^{n} |e_i^{(1)}|^p + (1-\alpha) \sum_{i=1}^{n} |e_i^{(2)}|^p$$

$$= \alpha k(\underline{e}^{(1)}) + (1-\alpha)k(\underline{e}^{(2)}) .$$

Hence, by definition, k is strictly convex. □

Theorem 2.2.2. Let $\underline{e} = \underline{y} - X\underline{\beta}$. Let $F(\underline{\beta}) = \sum_{i=1}^{n} |e_i|^p$, $p > 1$. Then, $F(\underline{\beta})$ is a convex function of $\underline{\beta}$. (It holds also for the case of $n < m$.) Further, $F(\underline{\beta})$ is a strictly convex function of $\underline{\beta}$ when X is full-rank.

Proof: Let $\underline{\beta}^{(1)} \in \mathbb{R}^m$, $\underline{\beta}^{(2)} \in \mathbb{R}^m$ such that $\underline{\beta}^{(1)} \neq \underline{\beta}^{(2)}$. Let $\underline{e}^{(1)} = \underline{y} - X\underline{\beta}^{(1)}$ and $\underline{e}^{(2)} = \underline{y} - X\underline{\beta}^{(2)}$. Let $0 < \alpha < 1$. Then,

$F(\alpha\underline{\beta}^{(1)}+(1-\alpha)\underline{\beta}^{(2)}) = k(\underline{y}-X(\alpha\underline{\beta}^{(1)}+(1-\alpha)\underline{\beta}^{(2)}))$, where k is as in Corollary 2.2.2,

$$= k(\alpha(\underline{y} - X\underline{\beta}^{(1)}) + (1-\alpha)(\underline{y} - X\underline{\beta}^{(2)}))$$

$$= k(\alpha\underline{e}^{(1)} + (1-\alpha)\underline{e}^{(2)}) .$$

If $\underline{e}^{(1)} \neq \underline{e}^{(2)}$, by Corollary 2.2.2,

$$k(\alpha\underline{e}^{(1)} + (1-\alpha)\underline{e}^{(2)}) < \alpha\, k(\underline{e}^{(1)}) + (1-\alpha)k(\underline{e}^{(2)})$$

i.e., $$F(\alpha\underline{\beta}^{(1)} + (1-\alpha)\underline{\beta}^{(2)}) < \alpha\, F(\underline{\beta}^{(1)}) + (1-\alpha)F(\underline{\beta}^{(2)}) .$$

If $\underline{e}^{(1)} = \underline{e}^{(2)}$, then

$$k(\alpha\underline{e}^{(1)} + (1-\alpha)\underline{e}^{(2)}) = k(\underline{e}^{(1)})$$

$$= k(\underline{e}^{(2)})$$

i.e., $$F(\alpha\underline{\beta}^{(1)} + (1-\alpha)\underline{\beta}^{(2)}) = F(\underline{\beta}^{(1)})$$

$$= F(\underline{\beta}^{(2)})$$

$$= \alpha F(\underline{\beta}^{(1)}) + (1-\alpha)F(\underline{\beta}^{(2)}) .$$

Thus, by definition, F is convex.

When X is full-rank, suppose $\underline{e}^{(1)} = \underline{e}^{(2)}$, then $X\underline{\beta}^{(1)} = X\underline{\beta}^{(2)}$, implies $\underline{\beta}^{(1)} = \underline{\beta}^{(2)}$, which is a contradiction. Hence, $\underline{e}^{(1)} \neq \underline{e}^{(2)}$. Therefore, we get only inequality, i.e.,

$$F(\alpha\underline{\beta}^{(1)} + (1-\alpha)\underline{\beta}^{(2)}) < \alpha\, F(\underline{\beta}^{(1)}) + (1-\alpha)F(\underline{\beta}^{(2)}) .$$

Thus, by definition, F is strictly convex when X is full-rank. □

2.3 Existence and Uniqueness of ℓ_p Estimates

Barrodale and Roberts (1970) expressed an ℓ_p estimation problem with full-rank X as a nonlinear programming problem with linear constraints and demonstrated the existence of ℓ_p estimates. Moreover,

they indicated that there is a unique ℓ_p estimate when $p > 1$ and ℓ_p estimates form a convex subset when $p = 1$. We will follow their approach in our discussion in this section.

We first state an elementary result in mathematical analysis.

Lemma 2.3.1. Let $h(\underline{\beta})$ be a real-valued continuous function defined on a subset A in $\mathbb{R}^k$. Let B be a closed and bounded subset in A. Then, there exists a $\underline{\beta}^* \in B$, such that $h(\underline{\beta}^*) \leq h(\underline{\beta})$ for all $\underline{\beta} \in B$.

By the approach of Barrodale and Roberts (1970), we can express the ℓ_p estimation problem in Chapter 1 as:

$$\text{minimize } \sum_{i=1}^{n} (u_i^p + v_i^p) \quad \text{such that}$$

$$y_i = \sum_{j=1}^{m} (b_j - c_j)x_{ij} + u_i - v_i, \quad 1 \leq i \leq n$$

where all b_j, c_j, u_i, and v_i are nonnegative. Let $\underline{b} = (b_1 b_2 \ldots b_m)^T$, $\underline{c} = (c_1 c_2 \ldots c_m)^T$, $\underline{u} = (u_1 u_2 \ldots u_n)^T$, and $\underline{v} = (v_1 v_2 \ldots v_n)^T$. Let $A = \{(\underline{b}, \underline{c}, \underline{u}, \underline{v}) \in \mathbb{R}^{2m+2n} | (\underline{b}, \underline{c}, \underline{u}, \underline{v}) \geq \underline{0} \text{ and } \underline{y} = X(\underline{b} - \underline{c}) + \underline{u} - \underline{v}\}$ Note that A is a closed convex subset in $\mathbb{R}^{2m+2n}$. Let h be a real-valued function defined on A, such that $h(\underline{b}, \underline{c}, \underline{u}, \underline{v}) = \sum_{i=1}^{n} u_i^p + v_i^p$. Then, we want to find a $(\underline{b}^*, \underline{c}^*, \underline{u}^*, \underline{v}^*) \in A$, such that $h(\underline{b}^*, \underline{c}^*, \underline{u}^*, \underline{v}^*) \leq h(\underline{b}, \underline{c}, \underline{u}, \underline{v})$ for all $(\underline{b}, \underline{c}, \underline{u}, \underline{v}) \in A$.

Let the subset $B = \{(\underline{b}, \underline{c}, \underline{u}, \underline{v}) \in \mathbb{R}^{2m+2n} | u_i \leq \rho, \ v_i \leq \rho, \ 1 \leq i \leq n\}$, where $\rho = \left(\sum_{i=1}^{n} |y_i|^p \right)^{\frac{1}{p}}$. Note that B is a closed

subset in $\mathbb{R}^{2m+2n}$. Since A and B are closed, $A \cap B$ is closed. Since X is full-rank, there exists a nonsingular X_1 such that $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ after a proper row arrangement. Let $\underline{y} = \begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \end{pmatrix}$, $\underline{u} = \begin{pmatrix} \underline{u}_1 \\ \underline{u}_2 \end{pmatrix}$, and $\underline{v} = \begin{pmatrix} \underline{v}_1 \\ \underline{v}_2 \end{pmatrix}$ accordingly. Then, given $\underline{u}$ and $\underline{v}$, we can solve $\underline{y}_1 = X_1(\underline{b} - \underline{c}) + \underline{u}_1 - \underline{v}_1$ for $\underline{b} - \underline{c}$. Since u_i, v_i, $1 \leq i \leq n$ are bounded in B , $b_j - c_j$, $1 \leq j \leq m$, are bounded in $A \cap B$. Let $C = \{(\underline{b}, \underline{c}, \underline{u}, \underline{v}) \in \mathbb{R}^{2m+2n} | b_j c_j = 0 , 1 \leq j \leq m , \text{ and } u_i v_i = 0 , 1 \leq i \leq n\}$. Note that C is a closed subset in $\mathbb{R}^{2m+2n}$. Now, $A \cap B \cap C$ is closed. $A \cap B \cap C$ is not empty since it contains the vector $(\underline{b}^o, \underline{c}^o, \underline{u}^o, \underline{v}^o)$, where $\underline{b}^o = \underline{c}^o = \underline{0}$ and $u_i^o v_i^o = 0$, $1 \leq i \leq n$. Also, b_j, c_j, $1 \leq j \leq m$, are bounded, as u_i, v_i, $1 \leq i \leq n$, are bounded in $A \cap B \cap C$. By Lemma 2.3.1, there exists $(\underline{b}^*, \underline{c}^*, \underline{u}^*, \underline{v}^*) \in A \cap B \cap C$, such that $h(\underline{b}^*, \underline{c}^*, \underline{u}^*, \underline{v}^*) \leq h(\underline{b}, \underline{c}, \underline{u}, \underline{v})$ for all $(\underline{b}, \underline{c}, \underline{u}, \underline{v}) \in A \cap B \cap C$. Let $(\underline{b}', \underline{c}', \underline{u}', \underline{v}') \in A$ such that $h(\underline{b}', \underline{c}', \underline{u}', \underline{v}') \leq h(\underline{b}, \underline{c}, \underline{u}, \underline{v})$ for all $(\underline{b}, \underline{c}, \underline{u}, \underline{v}) \in A$. Suppose $u_i' > \rho$ or $v_i' > \rho$ for some $1 \leq i \leq n$, then $\sum_{i=1}^{n} (u_i'^p + v_i'^p) > \sum_{i=1}^{n} |y_i|^p$, i.e., $h(\underline{b}', \underline{c}', \underline{u}', \underline{v}') > h(\underline{b}^o, \underline{c}^o, \underline{u}^o, \underline{v}^o)$, which is a contradiction. Therefore, $u_i' \leq \rho$, $v_i' \leq \rho$, $1 \leq i \leq n$. Also, we would choose $(\underline{b}', \underline{c}', \underline{u}', \underline{v}')$ such that $u_i' v_i' = 0$, $1 \leq i \leq n$, and $b_j' c_j' = 0$, $1 \leq j \leq m$. Thus, $(\underline{b}', \underline{c}', \underline{u}', \underline{v}') \in A \cap B \cap C$. Hence, $(\underline{b}', \underline{c}', \underline{u}', \underline{v}') = (\underline{b}^*, \underline{c}^*, \underline{u}^*, \underline{v}^*)$. Thus, we can derive the following Lemma.

Lemma 2.3.2. There always exists an ℓ_p estimate regardless of the rank of X.

Proof: We just demonstrated the case when X is full-rank. If X is not full-rank, as indicated in Section 2.1, we can work with a full-rank linear model and find an ℓ_p estimate in the full-rank linear model. Then, we can retrieve an ℓ_p estimate in the original linear model as indicated in Section 2.1. □

Next we will prove the existence of a constrained ℓ_p estimate with full-rank X. Let the constrained ℓ_p estimation problem be expressed as:

$$\text{minimize } h(\underline{b}, \underline{c}, \underline{u}, \underline{v}) \text{ such that } (\underline{b}, \underline{c}, \underline{u}, \underline{v}) \in A \cap D$$

where D is a closed subset in $\mathbb{R}^{2m+2n}$. Since we can always choose an optimal vector such that $b_j c_j = 0$, $1 \leq j \leq m$, and $u_i v_i = 0$, $1 \leq i \leq n$, i.e., there is always an optimal vector in $A \cap C \cap D$. Hence, we can consider the equivalent problem:

$$\text{minimize } h(\underline{b}, \underline{c}, \underline{u}, \underline{v}) \text{ such that } (\underline{b}, \underline{c}, \underline{u}, \underline{v}) \in A \cap C \cap D .$$

Let $(\tilde{\underline{b}}, \tilde{\underline{c}}, \tilde{\underline{u}}, \tilde{\underline{v}}) \in A \cap C \cap D$. Let $\tilde{\rho} = (h(\tilde{\underline{b}}, \tilde{\underline{c}}, \tilde{\underline{u}}, \tilde{\underline{v}}))^{\frac{1}{p}} = \left(\sum_{i=1}^{n} (\tilde{u}_i^p + \tilde{v}_i^p) \right)^{\frac{1}{p}}$. Let $B' = \{(\underline{b}, \underline{c}, \underline{u}, \underline{v}) \mid u_i \leq \tilde{\rho}, v_i \leq \tilde{\rho}, 1 \leq i \leq n\}$.

Note that $A \cap B' \cap C \cap D$ is closed and is not empty, since $\tilde{u}_i \leq \tilde{\rho}$, $\tilde{v}_i \leq \tilde{\rho}$, $1 \leq i \leq n$, implies that $(\tilde{\underline{b}}, \tilde{\underline{c}}, \tilde{\underline{u}}, \tilde{\underline{v}}) \in A \cap B' \cap C \cap D$. Also, $A \cap B' \cap C$ is bounded as discussed in the unconstrained case, hence, $A \cap B' \cap C \cap D$ is bounded. Then, by Lemma 2.3.1, there exists a locally

optimal vector for all $(\underline{b}, \underline{c}, \underline{u}, \underline{v}) \in A \cap B' \cap C \cap D$. Suppose that $(\underline{b}'', \underline{c}'', \underline{u}'', \underline{v}'') \in A \cap C \cap D$ such that $h(\underline{b}'', \underline{c}'', \underline{u}'', \underline{v}'') \leq h(\underline{b}, \underline{c}, \underline{u}, \underline{v})$ for all $(\underline{b}, \underline{c}, \underline{u}, \underline{v}) \in A \cap C \cap D$ and $u_i'' > \tilde{\rho}$ or $v_i'' > \tilde{\rho}$ for some $i \leq i \leq n$. Then, $\sum_{i=1}^{n} (u_i''^p + v_i''^p) > \tilde{\rho}^p = \sum_{i=1}^{n} (\tilde{u}_i^p + \tilde{v}_i^p)$, i.e., $h(\underline{b}'', \underline{c}'', \underline{u}'', \underline{v}'') > h(\tilde{\underline{b}}, \tilde{\underline{c}}, \tilde{\underline{u}}, \tilde{\underline{v}})$, which is a contradiction. Thus, $u_i'' \leq \tilde{\rho}$, $v_i'' \leq \tilde{\rho}$, $1 \leq i \leq n$, i.e., $(\underline{b}'', \underline{c}'', \underline{u}'', \underline{v}'') \in A \cap B' \cap C \cap D$. Therefore, the locally optimal vector is indeed a global optimal vector. Thus, we have the following Lemma.

Lemma 2.3.3. If the region of constraints is nonempty and closed, then there exists a constrained ℓ_p estimate with full-rank X.

For the uniqueness of ℓ_p estimate, as indicated in Section 2.1, the set of ℓ_p estimate forms a hypersubspace in R^m when X is not full-rank. Hence, ℓ_p estimate is not unique when X is not full-rank. In the rest of the section, we will discuss the uniqueness of ℓ_p estimate when X is full-rank.

The uniqueness of ℓ_p estimate is also clear by expressing ℓ_p estimation problem as the nonlinear programming problem. Note that h is linear when $p = 1$ and strictly convex when $p > 1$. Hence, when $p = 1$, there are infinite number of least absolute estimates if there are more than one. Moreover, the set of least absolute estimates is convex. When $p > 1$, suppose that $\underline{\beta}^1$ and $\underline{\beta}^2$ are ℓ_p estimates, then $h(\underline{\beta}_1) = h(\underline{\beta}_2)$, and $h(\alpha\underline{\beta}_1 + (1-\alpha)\underline{\beta}_2) < \alpha\, h(\underline{\beta}_1) + (1-\alpha)h(\underline{\beta}_2)$, for some $0 < \alpha < 1$. Hence, $h(\alpha\underline{\beta}_1 + (1-\alpha)\underline{\beta}_2) < h(\underline{\beta}_1)$,

which is a contradiction since $\underline{\beta}_1$ is an ℓ_p estimate. Therefore, there is a unique ℓ_p estimate when X is full-rank and $p > 1$.

3. DESCENT METHODS FOR ℓ_1 ESTIMATION

Usow (1967) discussed a geometric problem equivalent to the ℓ_1 estimation problem. He found that there exists an ℓ_1 estimate on one of the vertices of a convex polytope defined by the intersection of some 2^n half-spaces in R^{m+1}, and he proposed a descent method to find a lowest vertex, which is an ℓ_1 estimate. Bloomfield and Steiger (1980) proposed a method for ℓ_1 estimation which is related to the descent technique of Usow (1967). They claimed that their method is faster than the simplified simplex method of Barrodale and Roberts (1973) when the number of data points is large. Note that the method of Barrodale and Roberts (1973) is generally thought to be the most efficient method for ℓ_1 estimation today. Hence, there is at this time an open question as to which method is preferable. We will not attempt to resolve this question in this chapter.

The set of rows in the design matrix X of dimension $n \times m$, $n > m$, is Chebyshev if any m rows in X are linearly independent. Abdelmalek (1971) commented on the restriction of the Chebyshev condition which Usow (1967) had imposed on the problem. In this chapter, we will utilize some results from Bloomfield and Steiger (1980) to obtain the same results as those of Usow (1967) without the Chebyshev condition. Also, Lemma 4.4 of Usow (1967), which provides a proof of convergence of his method, is not correct. We have modified the lemma and will provide a proof without the Chebyshev condition. Then, we can obtain a proof of convergence of the method of Bloomfield and Steiger (1980).

Note that Bloomfield and Steiger (1980) did not explicitly prove convergence of their method.

3.1 Results due to Bloomfield and Steiger (1980)

In this section, some results from Bloomfield and Steiger (1980) with emphasis on identifying some independent rows in X at which residuals vanish will be discussed. Also, the specific condition under which their method is allowed to terminate will be described. The following two important results, called Result A and Result B, are due to Bloomfield and Steiger (1980).

Result A: When $m = 1$, the objective function is

$$F(\beta) = \sum_{i=1}^{n} |y_i - x_i\beta| = \sum_{i=1}^{n} |x_i| \left|\frac{y_i}{x_i} - \beta\right| .$$

The minimizing value of β is thus the weighted median of the ratio $\frac{y_i}{x_i}$, with respect to weights $|x_i| \neq 0$. This weighted median may be defined as any value $\hat{\beta}$ such that

$$\sum_{i:\, \frac{y_i}{x_i} = \hat{\beta}} |x_i| \geq \left| \sum_{i:\, \frac{y_i}{x_i} < \hat{\beta}} |x_i| - \sum_{i:\, \frac{y_i}{x_i} > \hat{\beta}} |x_i| \right| .$$

Say $\hat{\beta} = \frac{y_q}{x_q}$, then the q^{th} term of the sum $F(\hat{\beta})$ is zero.

Result B: There is an ℓ_1 estimate $\underline{\beta}^*$ for which at least t of the residuals vanish, t being the rank of X, and the corresponding rows in X are linearly independent. We will now present a proof of this

result with emphasis on the linearly independence of t rows in X.

Proof: Let $F(\underline{\beta}) = \sum_{i=1}^{n} |y_i - \underline{x}_i^T\underline{\beta}|$, where $\underline{x}_i$ is the i^{th} row of X. Let $\underline{\beta}^o$ be an ℓ_1 estimate, i.e.,

$$F(\underline{\beta}^o) \leq F(\underline{\beta}) \quad \text{for all} \quad \underline{\beta} \in \mathbb{R}^m .$$

We suppose that $y_i - \underline{x}_i^T\underline{\beta}^o = 0$ for $i = i_1, i_2, \ldots, i_r$, $\underline{x}_{i_1}^T, \underline{x}_{i_2}^T, \ldots, \underline{x}_{i_r}^T$ are linearly independent, and $0 \leq r < t$. Let K_1 be the row space spanned by $\underline{x}_{i_1}^T, \underline{x}_{i_2}^T, \ldots, \underline{x}_{i_r}^T$. Since $r < t$, there is a row, say $\underline{x}_u^T$, not in K_1. Let K be the row space spanned by K_1 and $\underline{x}_u^T$. Let $\{\alpha\underline{d} | \alpha \in \mathbb{R}, \underline{d} \in \mathbb{R}^m\}$ be the orthogonal complement of K_1 in K. Thus, $\underline{x}_i^T\underline{d} = 0$ for $i = i_1, i_2, \ldots, i_r$. Note that $\underline{x}_u^T\underline{d} \neq 0$, since otherwise $\underline{x}_u^T \in K_1$ which is a contradiction. Now, the function

$$S(\theta) = F(\underline{\beta}^o + \theta\underline{d}) = \sum_{i=1}^{n} |y_i - \underline{x}_i^T(\underline{\beta}^o + \theta\underline{d})|$$

is a sum with zero terms when $i = i_1, i_2, \ldots, i_r$ for all θ and $F(\underline{\beta}^o) = S(0)$. If we write

$$S(\theta) = \sum_{i=1}^{n} |r_i - \theta w_i| ,$$

where $r_i = y_i - \underline{x}_i^T\underline{\beta}^o$ and $w_i = \underline{x}_i^T\underline{d}$, by Result A the minimizing value of θ is the weighted median of the ratio $\frac{r_i}{w_i}$ with respect to weights $|w_i| \neq 0$, say $\hat{\theta} = \frac{r_q}{w_q}$, then the q^{th} term of the sum $S(\hat{\theta})$ is zero. Note that $w_i = 0$ for $i = i_1, i_2, \ldots, i_r$, hence, $q \neq i_k$ for $1 \leq k \leq r$. Also, $w_u \neq 0$ assures there is such $\frac{r_q}{w_q}$. Thus, $F(\underline{\beta}^o + \hat{\theta}\underline{d})$ has $r + 1$

zero residuals at $i = i_1, \ldots, i_m$ or $i = q$, and $F(\underline{\beta}^o + \hat{\theta}\underline{d}) = S(\hat{\theta}) \leq S(0) = F(\underline{\beta}^o) \leq F(\underline{\beta})$, for all $\underline{\beta} \in R^m$. Moreover, $w_q = \underline{x}_q^T\underline{d} \neq 0$ implies $\underline{x}_q^T \notin K_1$, hence, implies that $\{\underline{x}_{i_1}^T, \underline{x}_{i_2}^T, \ldots, \underline{x}_{i_r}^T, \underline{x}_q^T\}$ are linearly independent.

This argument holds until r is incremented to t. Then, we have an ℓ_1 estimate $\underline{\beta}^*$ such that $F(\underline{\beta}^*)$ has t zero residuals and the corresponding rows in X are linearly independent. □

Now we assume that X is full-rank. By Result B, there is an ℓ_1 estimate at which residuals vanish at m linearly independent rows in X. Hence, we can concentrate on all subsets consisting of m linearly independent rows in X. Note that there is a finite number of such subsets. We then compute $\underline{\beta}$ for which residuals vanish at each subset. One of these $\underline{\beta}$'s with the least $F(\underline{\beta})$ value will be identified as an ℓ_1 estimate. Bloomfield and Steiger (1980) proposed a method which requires computation of only a small portion of $\underline{\beta}$'s discussed above. The method involves two stages for each iteration. The problem in the first stage is to find a new row to replace a designated row in X which is thought to be the most desirable row to be deleted for the current iteration. We can start with any m linearly independent rows in X.

(i) Replacement stage: Let the current m linearly independent rows be $\underline{x}_1^T, \underline{x}_2^T, \ldots, \underline{x}_m^T$ and $\underline{x}_m^T$ be the designated row to be replaced. Let

$$\underline{x}_i^T\underline{\beta}^o = y_i \ , \quad 1 \leq i \leq m$$

$$\underline{d} \neq 0 \ \text{ such that } \ \underline{x}_i^T\underline{d} = 0 \ , \quad 1 \leq i \leq m-1$$

and $$S(\theta) = F(\underline{\beta}^o + \theta\underline{d}) = \sum_{i=1}^{n} |y_i - \underline{x}_i^T(\underline{\beta}^o + \theta\underline{d})| .$$

For every θ, $S(\theta)$ is a sum with zero terms when $1 \le i \le m-1$. Let us rewrite

$$S(\theta) = \sum_{i=1}^{n} |r_i - \theta w_i|$$

where $r_i = y_i - \underline{x}_i^T\underline{\beta}^o$ and $w_i = \underline{x}_i^T\underline{d}$. By Result A, $S(\theta)$ is minimized at $\theta = \hat{\theta}$, the weighted median of $\frac{r_i}{w_i}$ with respect to weights $|w_i| \neq 0$, say $\hat{\theta} = \frac{r_q}{w_q}$. Note that $w_m = \underline{x}_m^T\underline{d} \neq 0$ assures the existence of such $\frac{r_q}{w_q}$. Then $S(\hat{\theta})$ is a sum of zero terms when $1 \le i \le m-1$ or $i = q$. Also,

$$F(\underline{\beta}^o + \hat{\theta}\underline{d}) = S(\hat{\theta}) \le S(0) = F(\underline{\beta}^o) .$$

Moreover, $w_q = \underline{x}_q^T\underline{d} \neq 0$ implies $\underline{x}_q^T$ is not in the subspace spanned by $\underline{x}_1^T, \underline{x}_2^T, \ldots, \underline{x}_{m-1}^T$, hence, $\underline{x}_1^T, \underline{x}_2^T, \ldots, \underline{x}_{m-1}^T, \underline{x}_q^T$ are linearly independent.

Thus, by the replacement technique we can proceed to another m linearly independent rows $\underline{x}_1^T, \underline{x}_2^T, \ldots, \underline{x}_{m-1}^T, \underline{x}_q^T$ such that residuals of $\underline{\beta}^o + \hat{\theta}\underline{d}$ vanish at the m rows and the objective function is reduced at $\underline{\beta}^o + \hat{\theta}\underline{d}$. The next problem that we have is deciding which row to replace at each stage. In other words, we need a criterion for deciding on a "good" row to delete from the current subset so that it can be replaced according to the replacement technique described above. The Bloomfield and Steiger (1980) criterion for deletion is as follows. Note that it is based on the gradient.

(ii) Deletion stage: Note that $S(\theta) = \sum_{i=1}^{n} |r_i - \theta w_i|$ is a convex, piecewise linear function of θ, where $r_i = y_i - \underline{x}_i^T \beta^o$ and $w_i = \underline{x}_i^T \underline{d}$.

Let

$$S(\theta) = \sum_{i=1}^{n} |w_i| \left|\frac{r_i}{w_i} - \theta\right| = \sum_{i=1}^{n} |w_i| |v_i - \theta|$$

where $v_i = \frac{r_i}{w_i}$ and $w_i \neq 0$. The left-hand derivative at $\theta=0$ is

$$\frac{\partial}{\partial \theta} S(\theta)\Big|_{\theta=0-} = \sum_{i=1}^{n} |w_i| \operatorname{sign}(v_i - \theta)(-1)\Big|_{\theta=0-}$$

$$= -\sum_{v_i>0} |w_i| + \sum_{v_i<0} |w_i| - \sum_{v_i=0} |w_i| .$$

The right-hand derivative at $\theta=0$ is

$$\frac{\partial}{\partial \theta} S(\theta)\Big|_{\theta=0+} = -\sum_{v_i>0} |w_i| + \sum_{v_i<0} |w_i| + \sum_{v_i=0} |w_i| .$$

Thus, the larger of the left-hand derivative at 0 and the negative of the right-hand derivative at 0 is

$$\left| \sum_{v_i<0} |w_i| - \sum_{v_i>0} |w_i| \right| - \sum_{v_i=0} |w_i| .$$

Let

$$\rho = \frac{\left| \sum_{v_i<0} |w_i| - \sum_{v_i>0} |w_i| \right| - \sum_{v_i=0} |w_i|}{\sum |w_i|}$$

then ρ is independent of the magnitude of $\underline{d}$.

Let Z be the inverse matrix of $\begin{pmatrix} \underline{x}_1^T \\ \underline{x}_2^T \\ \vdots \\ \underline{x}_m^T \end{pmatrix}$. Let $\underline{z}_j$ be the j^{th} column in Z. Then $\underline{x}_i'\underline{z}_j = 0$, for $i \neq j$. Thus, any scalar multiple of $\underline{z}_j$ is a proper $\underline{d}$ in the replacement stage when row $\underline{x}_j$ has been identified for replacement. Hence, we can use columns in Z to compute ρ's, choose one with the largest positive ρ, and delete the corresponding row in $\begin{pmatrix} \underline{x}_1^T \\ \underline{x}_2^T \\ \vdots \\ \underline{x}_m^T \end{pmatrix}$.

If $\rho \leq 0$ for all columns in Z, we should start over with another m linear independent rows in X such that residuals at $\underline{\beta}^o$ vanish in the m rows. If $\rho \leq 0$ for all such m rows in X. Then $\underline{\beta}^o$ is an ℓ_1 estimate. (A proof will be given in the next section.)

3.2 Discussion of the Usow's Method

We will next describe a similar ℓ_1 approximation problem and notation which is contained in the paper of Usow (1967) and will give a counterexample to show that Lemma 4.4 of Usow (1967) is incorrect. Then, using results in Section 3.1, and Lemma 4.3 of Usow (1967), we can show that the Chebyshev condition is not necessary. Hence, we can modify Usow's Lemma 4.4 and present a proof without the

Chebyshev condition. The modified Lemma then can serve as a proof of convergence of Usow's method. Furthermore, the fact that the method of Bloomfield and Steiger (1980) is convergent follows immediately from this result. Finally, a modified algorithm for Usow's method will be presented. In the remainder of this chapter, we assume that the design matrix X is full-rank.

Usow (1967) considered the problem of finding $\underline{A}^*$ which minimizes

$$R(\underline{A}) = \sum_{i=1}^{n} |L(\underline{A},x_i) - f(x_i)|$$

where $\underline{A}$ denotes the parameter vector $(a_1, a_2, \ldots, a_m)^T$, $L(\underline{A},x) = \sum_{i=1}^{m} a_i\phi_i(x)$, and $\phi_1(x), \phi_2(x), \ldots, \phi_m(x), f(x)$ are real valued functions defined on $X = \{x_1, x_2, \ldots, x_n\}$. $L(\underline{A},x)$ in a "Lagrangian" form of interpolating $f(x)$ at the points $U = \{u_1, u_2, \ldots, u_m\}$ is defined as follows. Let

$$L(\underline{A},x) = \sum_{i=1}^{m} \tilde{a}_i\pi_i(x)$$

where $\tilde{a}_i = f(u_i)$, $\pi_j(u_i) = \delta_{ij}$, $i,j = 1, 2, \ldots, m$ and δ_{ij} denotes the Kronecker delta. Let

$$\pi_j(x) = \sum_{k=1}^{m} b_k^j\phi_k(x)$$

and let

$$\pi_j(u_i) = \sum_{k=1}^{m} b_k^j\phi_k(u_i) \quad , \quad i,j = 1, 2, \ldots, m$$

be expressed in matrix form, then,

$$\begin{pmatrix} \phi_1(u_1) & \phi_2(u_1) & \cdots & \phi_m(u_1) \\ \phi_1(u_2) & \phi_2(u_2) & \cdots & \phi_m(u_2) \\ \vdots & \vdots & & \vdots \\ \phi_1(u_m) & \phi_2(u_m) & \cdots & \phi_m(u_m) \end{pmatrix} \begin{pmatrix} b_1^1 & b_1^2 & \cdots & b_1^m \\ b_2^1 & b_2^2 & \cdots & b_2^m \\ \vdots & \vdots & & \vdots \\ b_m^1 & b_m^2 & \cdots & b_m^m \end{pmatrix} = I \ .$$

Thus, $[b_i^j] = [\phi_j(u_i)]^{-1}$ if the m rows in $[\phi_j(u_i)]$ are linearly independent. Let $\underline{b}_j = [b_1^j, b_2^j, \ldots, b_m^j]^T$, $1 \leq j \leq m$. Then $\underline{b}_j$ is $\underline{z}_j$ in Section 3.1, the j^{th} column of $[\phi_j(u_i)]^{-1}$. Thus, the "Lagrangian" form $L(\underline{A},x)$ is well-defined when residuals of $\underline{A}$ vanish at m linearly independent rows, i.e., for $x \in U = \{u_1, u_2, \ldots, u_m\}$. In the rest of the chapter we refer to a "Lagrangian" form only when it is well-defined.

Let $Z(\underline{A}) = \{x \in X | L(\underline{A},x) - f(x) = 0\}$ and let $\mu(Z(\underline{A}))$ denote the number of points in $Z(\underline{A})$. When $\mu(Z(\underline{A})) \geq m$, if we use another set U' of m points in $Z(\underline{A})$ such that the corresponding m rows in $[\phi_j(x_i)]$ are linearly independent, we get a distinct "Lagrangian" form $L'(\underline{A},x)$ at $\underline{A}$. For example let

$$X = \{x_1, x_2, x_3, x_4\} \ ,$$

$$[\phi_j(x_i)] = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 0 & 1 \\ 1 & -1/2 \end{pmatrix}$$ (note that the set of rows is Chebyshev),

and

$$[f(x_i)] = \begin{pmatrix} -1 \\ -2 \\ -1 \\ -1 \end{pmatrix} .$$

Let $\underline{A}_o = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$. Then $Z(\underline{A}_o) = \{x_1, x_2, x_3\}$. Let $U = \{x_1, x_3\}$ and $U' = \{x_2, x_3\}$. Let the "Lagrangian" form $L(\underline{A}_o, x)$ interpolate $f(x)$ at U. Thus,

$$L(\underline{A}_o, x) = (-1)\,\pi_1(x) + (-1)\,\pi_2(x)$$

where $\pi_j(x) = b_1^j \phi_1(x) + b_2^j \phi_2(x), \quad j = 1, 2$

and $$[b_i^j] = \begin{pmatrix} \phi_1(x_1) & \phi_2(x_1) \\ \phi_1(x_3) & \phi_2(x_3) \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}.$$

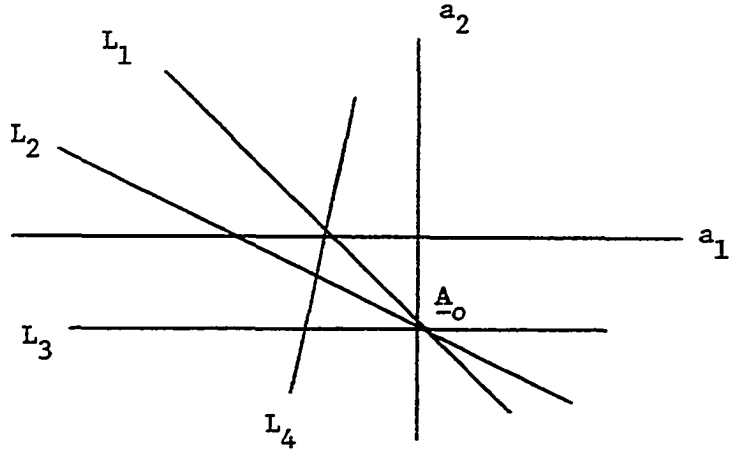
Let the "Lagrangian" form $L'(\underline{A}_o, x)$ interpolate $f(x)$ at U'. Thus,

$$L'(\underline{A}_o, x) = (-2)\,\pi_1'(x) + (-1)\,\pi_2'(x)$$

where $\pi_j'(x) = b_1'^j \phi_1(x) + b_2'^j \phi_2(x), \quad j = 1, 2,$

and $$[b_i'^j] = \begin{pmatrix} \phi_1(x_2) & \phi_2(x_2) \\ \phi_1(x_3) & \phi_2(x_3) \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}.$$

We can depict graphically, in the parameter space $\mathbb{R}^2$, for this given example, as follows,



where $L_i = \{(a_1, a_2) \mid \phi_1(x_i)a_1 + \phi_2(x_i)a_2 = f(x_i)\}$. Note that

$$\phi_1(x_3)b_1^1 + \phi_2(x_3)b_2^1 = 0 .$$

Hence, $$\phi_1(x_3)(a_1^o \pm \varepsilon\, b_1^1) + \phi_2(x_3)(a_2^o \pm \varepsilon\, b_2^1) = f(x_3) ,$$

i.e., $\underline{A}_o \pm \varepsilon\, \underline{b}^1$ is on the Line L_3, where $\underline{A}_o = \begin{pmatrix} a_1^o \\ a_2^o \end{pmatrix}$, $\underline{b}^1 = \begin{pmatrix} b_1^1 \\ b_2^1 \end{pmatrix}$,

and $\varepsilon > 0$. Let $\underline{b}^j = \begin{pmatrix} b_1^j \\ b_2^j \end{pmatrix}$ and $\underline{b}'^j = \begin{pmatrix} b_1'^j \\ b_2'^j \end{pmatrix}$. Similarly, we have $\underline{A}_o \pm \varepsilon\, \underline{b}^2$ on the Line L_1, $\underline{A}_o \pm \varepsilon\, \underline{b}'^1$ on the Line L_3, and $\underline{A}_o \pm \varepsilon\, \underline{b}'^2$ on the Line L_2. Thus, using distinct "Lagrangian" forms at $\underline{A}_o$, we are able to determine directions to the adjacent intersections.

We now state Lemma 4.4 of Usow (1967) and show that the above example is a counterexample to the lemma. Usow's Lemma 4.4 is stated as: If $(\underline{A}_s, d_s)$ is a vertex such that $R(\underline{A}_s \pm \varepsilon\, \underline{b}^i) \geq d_s$ for every $\varepsilon > 0$ and $1 \leq i \leq m$, then $(\underline{A}_s, d_s)$ is a lowest vertex.

Note that the set of rows in $[\phi_j(x_i)]$ of the example is Chebyshev as required in the paper of Usow (1967). Let $(\underline{A}_o, R(\underline{A}_o))$ be the $(\underline{A}_s, d_s)$ in the lemma. Usow (1967) proved that $(\underline{A}_s, R(\underline{A}_s))$ is a vertex if $\mu(Z(\underline{A}_s)) \geq m$. Now $\mu(Z(\underline{A}_o)) = 3 > 2$, hence $(\underline{A}_o, R(\underline{A}_o))$ is a vertex. We have that

$$L(\underline{A}_o, x) = (-1)\phi_2(x)$$

and $$R(\underline{A}_o) = \sum_{i=1}^{4} |L(\underline{A}_o, x_i) - f(x_i)| = 3/2 .$$

For every $\varepsilon > 0$,

$$L(\underline{A}_o \pm \varepsilon\, \underline{b}^1, x) = \pm \varepsilon\, \phi_1(x) + (-1)\, \phi_2(x)$$

then $$R(\underline{A}_o \pm \varepsilon\, \underline{b}^1) = \sum_{i=1}^{4} |L(\underline{A}_o \pm \varepsilon\, \underline{b}^1, x_i) - f(x_i)|$$

$$= \varepsilon + \varepsilon + 3/2 \pm \varepsilon$$

$$> 3/2 = R(\underline{A}_o) .$$

For every $\varepsilon > 0$,

$$L(\underline{A}_o \pm \varepsilon\, \underline{b}^2, x) = \pm \varepsilon\, \phi_1(x) + (-1 \pm \varepsilon)\, \phi_2(x)$$

then, $$R(\underline{A}_o \pm \varepsilon\, \underline{b}^2) = \sum_{i=1}^{4} |L(\underline{A}_o \pm \varepsilon\, \underline{b}^2, x_i) - f(x_i)|$$

$$= 0 + \varepsilon + \varepsilon + 3/2 \mp 3\varepsilon/2$$

$$> 3/2 = R(\underline{A}_o) .$$

Since all conditions on Usow's Lemma 4.4 are satisfied, we have $(\underline{A}_o, R(A_o))$ is a lowest vertex. However,

$$R((-6/5, -2/5)) = \sum_{i=1}^{4} |(-6/5)\phi_1(x_i) + (-2/5)\phi_2(x_i) - f(x_i)|$$

$$= 3/5 + 0 + 3/5 + 0$$

$$= 6/5$$

$$< 3/2 = R(\underline{A}_o)$$

Hence, $(\underline{A}_o, R(\underline{A}_o))$, in fact, is not a lowest vertex.

Usow's Lemma 4.4 is correct if $\mu(Z(\underline{A}_s)) = m$, but in general $\mu(Z(\underline{A}_s)) = s \geq m$. Note that there are $\binom{s}{m}$ distinct "Lagrangian" forms at $\underline{A}_s$ if $\{\underline{\phi}(x) | x \in Z(\underline{A}_s)\}$ is Chebyshev and less than $\binom{s}{m}$ if it is

not Chebyshev, where $\underline{\phi}(x) = (\phi_1(x)\ \phi_2(x)\ \ldots\ \phi_m(x))^T$ the row in $[\phi_j(x_i)]$ for $x \in X$. We will now modify the lemma such that it can provide a proper criterion for convergence of the Usow's descent method. Moreover, we can derive a proof without the Chebyshev condition. In fact, the Chebyshev condition is not necessary in the method. We reason as follows.

As indicated in Section 3.1, we can concentrate only on vectors in the parameter space which have the "Lagrangian" forms. Hence, we start with any vector in the parameter space which has a "Lagrangian" form. By the following lemma, which plays the same role as of the replacement technique in Section 3.1, we can find a descent direction to another vector in the parameter space which has a "Lagrangian" form. We would repeatedly apply this lemma until a proper criterion for convergence is reached. (Details of the method will be provided in a modified algorithm later on.) Since we do not use the Chebyshev condition in the method, it is not necessary then. The Usow's Lemma 4.3 is stated as: Let $L(\underline{A}_k, x)$ be in a "Lagrangian" form of interpolating $f(x)$ at $U = \{u_1, u_2, \ldots, u_m\}$ with the associated $\underline{b}^1, \underline{b}^2, \ldots, \underline{b}^m$. If $R(\underline{A}_k - \varepsilon \underline{b}^r) < R(\underline{A}_k)$ for some $\varepsilon > 0$, then there is an Δa_r such that $R(\underline{A}_k - \Delta a_r \underline{b}^r) = \min_{\theta>0} R(\underline{A}_k - \theta \underline{b}^r)$. This holds also for the case of $\varepsilon < 0$ and $\theta < 0$. (Note that the Chebyshev condition is not needed in the proof of this lemma.)

In addition to the proof of the lemma from Usow (1967),

$$\sum_{i=1}^{m} \phi_i(u_k) b_i^r = 0 \quad \text{for all} \quad k \neq r \ ,$$

and $$\sum_{i=1}^{m} \phi_i(x_j) b_i^r \neq 0 \ ,$$

implies that the row x_j is not in the subspace spanned by the $m-1$ linearly independent rows $u_1, u_2, \ldots, u_{r-1}, u_{r+1}, \ldots, u_m$. Therefore, these m rows are linearly independent. Hence, the "Lagrangian" form $L(\underline{A}_k - \Delta a_r \underline{b}^r, x)$ of interpolating $f(x)$ at the set of points $\{u_1, u_2, \ldots, u_{r-1}, u_{r+1}, \ldots, u_m, x_j\}$ is well-defined.

Further, $$R(\underline{A}_k - \theta \underline{b}^r) = \sum_{i=1}^{n} \left| f(x_i) - \sum_{j=1}^{m} \phi_j(x_i)(a_j^k - \theta b_j^r) \right| \ ,$$

where $$\sum_{j=1}^{m} \phi_j(u_i) b_j^r = 0 \ , \quad i = 1, 2, \ldots, r-1, r+1, \ldots, m \ ,$$

and $$\underline{A}_k = (a_1^k, a_2^k, \ldots, a_m^k)^T \ .$$

By Result A in Section 3.1, $\min_{\substack{\theta>0 \\ \text{or } \theta<0}} R(\underline{A}_k - \theta \underline{b}^r) = R(\underline{A}_k - \hat{\theta} \underline{b}^r)$, where $\hat{\theta}$ is a weighted median of $-(f(x_i) - \sum_{j=1}^{m} \phi_j(x_i) a_j^k) / \sum_{j=1}^{m} \phi_j(x_i) b_j^r$ with weights $\left| \sum_{j=1}^{m} \phi_j(x_i) b_j^r \right| \neq 0$ for $1 \leq i \leq n$.

By Usow's Lemma 4.3, we can proceed from a vector in the parameter space which has a "Lagrangian" form to another vector in the parameter space which has a "Lagrangian" form such that the objective function is reduced. Since there are finite vectors in the parameter space, which have "Lagrangian" forms, the condition in Usow's

Lemma 4.3 must not be satisfied for all "Lagrangian" forms at some vector in the parameter space, say $\underline{A}_s$. That is to say there exists $\underline{A}_s$ in the parameter space such that $R(\underline{A}_s \pm \varepsilon \underline{b}^j) \geq R(\underline{A}_s)$ for all $\underline{b}^j$'s in all "Lagrangian" forms at $\underline{A}_s$ and $\varepsilon > 0$. A theorem and its corollary, which modifies Usow's Lemma 4.4, are provided to prove that $\underline{A}_s$ is an ℓ_1 estimate. Note that we do not use the Chebyshev condition in their proofs. We next proceed to state and prove some lemmas and corollaries which are needed in the proof of the theorem.

Let $H(x)$ be the hyperplane defined by

$$H(x) = \{\underline{A} \in \mathbb{R}^m | \underline{\phi}^T(x)\underline{A} = f(x)\} \text{ , where } x \in Z(\underline{A}_s) .$$

Let $H(Z(\underline{A}_s)) = \{H(x) | x \in Z(\underline{A}_s)\}$.

Lemma 3.2.1. Let $L(\underline{A}_s, x)$ be in a "Lagrangian" form of interpolating $f(x)$ at $U = \{u_1, u_2, \ldots, u_m\}$ with the associated $\underline{b}^j$'s . Then $\underline{A}_s \pm \varepsilon \underline{b}^j$ is on the intersection of $m-1$ linearly independent hyperplanes in $H(Z(\underline{A}_s))$, where $\varepsilon > 0$ and $1 \leq j \leq m$. It also holds for the reverse direction.

Proof: ($\Rightarrow$) $\underline{\phi}^T(u_i)(\underline{A}_s \pm \varepsilon \underline{b}^j) = \underline{\phi}^T(u_i)\underline{A}_s = f(u_i)$ for $1 \leq i \leq m$ and $i \neq j$. Then $\underline{A}_s \pm \varepsilon \underline{b}^j$ is on the hyperplane $H(u_i)$ for $1 \leq i \leq m$ and $i \neq j$. Hence, $\underline{A}_s \pm \varepsilon \underline{b}^j$ is at the intersection of $m-1$ linearly independent hyperplanes $H(u_i)$, $1 \leq i \leq m$ and $i \neq j$, in $H(Z(\underline{A}_s))$.

($\Leftarrow$) Let $\underline{A}_s + \underline{z} \in H(x_i)$, $i = 1, 2, \ldots, m-1$, $x_i \in Z(\underline{A}_s)$, and $\{\underline{\phi}(x_i) | i = 1, 2, \ldots, m-1\}$ are linearly independent, where $\underline{z} \neq \underline{0}$. Then

$$\underline{\phi}^T(x_i)(\underline{A}_s + \underline{z}) = f(x_i) \ , \quad i = 1, 2, \ldots, m\text{-}1$$

implies $\underline{\phi}^T(x_i)\underline{z} = 0 \ , \quad i = 1, 2, \ldots, m\text{-}1 \ .$

Since $\mu(Z(\underline{A}_s)) \geq m$ and m of the corresponding rows in $(\phi_j(x_i))$ are linearly independent, we can find one, say x_m, such that $x_m \in Z(\underline{A}_s)$ and $\{\underline{\phi}(x_i) \mid i = 1, 2, \ldots, m\}$ are linearly independent. Let $L'(\underline{A}_s, x)$ be the corresponding "Lagrangian" form with $\underline{b}'^j$, $1 \leq j \leq m$. Suppose $\underline{\phi}^T(x_m)\underline{z} = 0$ then $\underline{z} = \underline{0}$, which is a contradiction. Hence, $\underline{\phi}^T(x_m)\underline{z} \neq 0$ and let $\underline{b}'^m = \dfrac{\underline{z}}{\underline{\phi}^T(x_m)\underline{z}}$. Thus, $\underline{A}_s + \underline{z} = \underline{A}_s \pm \varepsilon \, \underline{b}'^m$, where $\varepsilon = |\underline{\phi}^T(x_m)\underline{z}|$. □

Note that there are at most $\binom{s}{m-1}$ directions issuing from $\underline{A}_s$ on which we can apply the Usow's Lemma 4.3 to obtain descent directions to vectors in the parameter space which have "Lagrangian" forms. The following lemma and corollary will show that the sign of nonzero residual of $\underline{A}_s$ does not change along all these directions if all vectors are restricted to a sufficiently small neighborhood of $\underline{A}_s$ in the parameter space.

Lemma 3.2.2. $\text{Sign}(L(\underline{A}_s \pm \varepsilon \, \underline{b}^j, x) - f(x)) = \text{sign}(L(\underline{A}_s, x) - f(x))$ for $1 \leq j \leq m$ when $x \in X - Z(\underline{A}_s)$ and ε is sufficiently small.

Proof: Let $m_k^j = \dfrac{L(\underline{A}_s, x_k) - f(x_k)}{\pi_j(x_k)}$, for some $x_k \in X - Z(\underline{A}_s)$, such that

$$|m_k^j| = \min \{ \frac{|L(\underline{A}_s,x) - f(x)|}{|\pi_j(x)|} \mid x \in X - Z(\underline{A}_s) \} .$$

Let $0 < \varepsilon < |m_k^j|$ for $1 \leq j \leq m$. For some $x \in X - Z(\underline{A}_s)$, if

$$L(\underline{A}_s \pm \varepsilon \underline{b}^j,x) \leq f(x) < L(\underline{A}_s,x)$$

then $$L(\underline{A}_s,x) \pm \varepsilon \pi_j(x) \leq f(x) < L(\underline{A}_s,x)$$

hence, $$|\frac{L(\underline{A}_s,x) - f(x)}{\pi_j(x)}| \leq \varepsilon < |m_k^j|$$

which is a contradiction. Similarly, for some $x \in X - Z(\underline{A}_s)$, if

$$L(\underline{A}_s \pm \varepsilon \underline{b}^j,x) \geq f(x) > L(\underline{A}_s,x)$$

then it leads to the same contradiction. Hence,

$$\text{sign}(L(\underline{A}_s \pm \varepsilon \underline{b}^j,x) - f(x)) = \text{sign}(L(\underline{A}_s,x) - f(x))$$

for $1 \leq j \leq m$ when $x \in X - Z(\underline{A}_s)$ and ε is sufficiently small. □

Corollary 3.2.1. $\text{Sign}(L(\underline{A}_s \pm \varepsilon \underline{b}^j,x) - f(x)) = \text{sign}(L(\underline{A}_s,x) - f(x))$ for all "Lagrangian" forms at $\underline{A}_s$, $j = 1, 2, \ldots, m$, when $x \in X - Z(\underline{A}_s)$ and ε is sufficiently small.

Proof: Since there are finite distinct "Lagrangian" forms at $\underline{A}_s$, there are finite m_k^j's in Lemma 3.2.2. Hence, if we choose $\varepsilon > 0$ such that ε is smaller than all such $|m_k^j|$'s, we can get the conclusion. □

Theorem 3.2.1. Let $L(\underline{A}_s, x)$ be in a "Lagrangian" form such that

$$R(\underline{A}_s \pm \varepsilon \underline{b}^j) \geq R(\underline{A}_s) \ , \quad j = 1, 2, \ldots, m \ , \text{ for all } \ \varepsilon > 0 \ ,$$

is a local minimum.

Proof: Let $H^+(x)$ denote the half-space $\{\underline{A} \in \mathbb{R}^m | \underline{\phi}^T(x)\ \underline{A} \geq f(x)\}$ and $H^-(x)$ denote the half-space $\{\underline{A} \in \mathbb{R}^m | \underline{\phi}^T(x)\ \underline{A} \leq f(x)\}$, where $x \in Z(\underline{A}_s)$. Now $P = \{H^{p_1}(x_1) \cap H^{p_2}(x_2) \cap \ldots \cap H^{p_s}(x_s) | p_i = i = 1, 2, \ldots, s\}$ partitions any neighborhood of $\underline{A}_s$ in the sense that each element in P contains the boundary. Let $T \in P$ and $T \neq \emptyset$, $T = H^{p_1}(x_1) \cap H^{p_2}(x_2) \cap \ldots \cap H^{p_s}(x_s)$. We rewrite T as $H^{\tilde{p}_1}(v_1) \cap H^{\tilde{p}_2}(v_2) \cap \ldots \cap H^{\tilde{p}_t}(v_t)$ such that none of $H^{\tilde{p}_i}(v_i)$ could be dropped from the expression of T, then $t \leq s$. Also, $t \geq m$ since there are m linearly independent hyperplanes in $H(Z(\underline{A}_s))$. Let $H(T) = \{H(v_j) | 1 \leq j \leq t\}$. Let $Z = \{\underline{A}_s + \underline{z} \in T | \ \|\underline{z}\|_2 < \delta$ and $\underline{A}_s + \underline{z}$ is on the intersection of some $m-1$ linearly independent hyperplanes in $H(T)\}$. Then

$$\left\{ \sum_{i=1}^{m} \lambda_i (\underline{A}_s + \underline{z}_i) \Big| \sum_{i=1}^{m} \lambda_i = 1 \ , \ \lambda_i \geq 0 \ , \text{ and } \underline{A}_s + \underline{z}_i \in Z \ , \right.$$
$$\left. \text{for } \ 1 \leq i \leq m \right\} \supset T \cap C(\underline{A}_s) \ , \text{ where } \ C(\underline{A}_s)$$

is a small open neighborhood of $\underline{A}_s$. That is, let $\underline{B} \in T \cap C(\underline{A}_s)$ then

$$\underline{B} = \sum_{i=1}^{m} \lambda_i (\underline{A}_s + \underline{z}_i) = \underline{A}_s + \sum_{i=1}^{m} \lambda_i \underline{z}_i$$

for some $\lambda_i \geq 0$ such that $\sum_{i=1}^{m} \lambda_i = 1$ and some $\underline{z}_i \in Z$, $1 \leq i \leq m$. By Lemma 3.2.1, $\underline{A}_s + \underline{z}_i = \underline{A}_s \pm \varepsilon(i)\underline{b}^{j(i)}$ for $\underline{b}^{j(i)}$ under the "Lagrangian" form $L^{(i)}(\underline{A}_s, x)$, where $\varepsilon(i) = |\underline{\phi}^T(x_k)\underline{z}_i|$ for some $x_k \in Z(\underline{A}_s)$. Let $M = \max_{1 \leq i \leq s} \|\phi^T(x_i)\|_2$. Then $\varepsilon(i) \leq M\|\underline{z}_i\|_2 \leq M\delta$, for $1 \leq i \leq m$. Since $\underline{A}_s \pm \varepsilon(i)\underline{b}^{j(i)} \in T$ for each $x_\ell \in Z(\underline{A}_s)$, $\underline{A}_s \pm \varepsilon(i)\underline{b}^{j(i)} \in H^{p_\ell}(x_\ell)$ for $1 \leq i \leq m$. In other words, for each $x_\ell \in Z(\underline{A}_s)$, $\operatorname{sign}(\underline{\phi}^T(x_\ell)(\pm \varepsilon(i))\underline{b}^{j(i)})$ is constant for $1 \leq i \leq m$ and $\underline{\phi}^T(x_\ell)(\pm \varepsilon(i))\underline{b}^{j(i)} \neq 0$.

$$\text{Now, } R(\underline{B}) = R(\underline{A}_s + \sum_{i=1}^{m} \lambda_i(\pm \varepsilon(i))\underline{b}^{j(i)})$$

$$= \sum_{x \in X} |\underline{\phi}^T(x)\underline{A}_s + \sum_{i=1}^{m} \lambda_i(\pm \varepsilon(i))\underline{\phi}^T(x)\underline{b}^{j(i)} - f(x)|$$

$$= \sum_{x \in Z(\underline{A}_s)} |\sum_{i=1}^{m} \lambda_i(\pm \varepsilon(i))\underline{\phi}^T(x)\underline{b}^{j(i)}|$$

$$+ \sum_{x \in X - Z(\underline{A}_s)} |\sum_{i=1}^{m} \lambda_i(\underline{\phi}^T(x)\underline{A}_s \pm \varepsilon(i)\,\underline{\phi}^T(x)\underline{b}^{j(i)} - f(x))|$$

$$= I + II .$$

$$I = \sum_{x \in Z(\underline{A}_s)} \sum_{i=1}^{m} \lambda_i |\pm \varepsilon(i)\underline{\phi}^T(x)\underline{b}^{j(i)}|$$

$$= \sum_{i=1}^{m} \lambda_i \sum_{x \in Z(\underline{A}_s)} |\pm \varepsilon(i)\underline{\phi}^T(x)\underline{b}^{j(i)}|$$

$$= \sum_{i=1}^{m} \lambda_i \sum_{x \in Z(\underline{A}_s)} |L^{(i)}(\underline{A}_s, x) \pm \varepsilon(i)\underline{\phi}^T(x)\underline{b}^{j(i)} - f(x)|$$

$$= \sum_{i=1}^{m} \lambda_i \sum_{x \in Z(\underline{A}_s)} |L^{(i)}(\underline{A}_s \pm \varepsilon(i)\underline{b}^{j(i)}, x) - f(x)| \ .$$

$$II = \sum_{x \in X - Z(\underline{A}_s)} \left| \sum_{i=1}^{m} \lambda_i (L^{(i)}(\underline{A}_s \pm \varepsilon(i)\underline{b}^{j(i)}, x) - f(x)) \right|$$

$$= \sum_{x \in X - Z(\underline{A}_s)} \sum_{i=1}^{m} \lambda_i |L^{(i)}(\underline{A}_s \pm \varepsilon(i)\underline{b}^{j(i)}, x) - f(x)| \ .$$

Since $\varepsilon(i) \leq M\delta$, the $\varepsilon(i)$'s are small if we choose δ sufficiently small. Then apply Corollary 3.2.1 to obtain

$$II = \sum_{i=1}^{m} \lambda_i \sum_{x \in X - Z(\underline{A}_s)} |L^{(i)}(\underline{A}_s \pm \varepsilon(i)\underline{b}^{j(i)}, x) - f(x)| \ .$$

Thus, $$R(\underline{B}) = \sum_{i=1}^{m} \lambda_i \sum_{x \in X} |L^{(i)}(\underline{A}_s \pm \varepsilon(i)\underline{b}^{j(i)}, x) - f(x)|$$

$$= \sum_{i=1}^{m} \lambda_i R(\underline{A}_s \pm \varepsilon(i)\underline{b}^{j(i)})$$

$$\geq \sum_{i=1}^{m} \lambda_i R(\underline{A}_s) \ , \text{ and by the assumption}$$

$$R(\underline{B}) = R(\underline{A}_s) \ .$$

Since P partitions the small open neighborhood of $\underline{A}_s$, $C(\underline{A}_s)$, let $\underline{B} \in C(\underline{A}_s)$. Then there is a $T \in P$ such that $B \in T \cap C(\underline{A}_s)$. Thus, $R(\underline{B}) \geq R(\underline{A}_s)$. Therefore, $R(\underline{A}_s)$ is a local minimum. □

Corollary 3.2.2. $R(\underline{A}_s)$ is a local mimimal implies that it is a global minimal, i.e., $\underline{A}_s$ is an ℓ_1 estimate.

Proof: By Result B in Section 3.1, we need only consider vectors in the parameter space which have "Lagrangian" forms. Let $\underline{A}_o$ be one of these vectors. Then there exists $0 \leq \alpha < 1$ such that $\alpha\underline{A}_s + (1-\alpha)\underline{A}_o \in C(\underline{A}_s)$, where $C(\underline{A}_s)$ is as in Theorem 3.2.1. Hence,

$$R(\alpha\underline{A}_s + (1-\alpha)\underline{A}_o) \geq R(\underline{A}_s) .$$

Suppose that $R(\underline{A}_o) < R(\underline{A}_s)$. By Theorem 2.2.1, $R(\underline{A})$ is convex. Then

$$R(\alpha\underline{A}_s + (1-\alpha)\underline{A}_o) \leq \alpha R(\underline{A}_s) + (1-\alpha)R(\underline{A}_o) < R(\underline{A}_s)$$

which is a contradiction. Thus, $R(\underline{A}_o) \geq R(\underline{A}_s)$. □

The following is a proof of convergence of the method of Bloomfield and Steiger (1980).

Corollary 3.2.3. In Section 3.1, if $\rho \leq 0$ for all columns in the inverse matrix of any m linearly independent rows in X at which residuals of $\underline{\beta}^o$ vanish, then $\underline{\beta}^o$ is an ℓ_1 estimate.

Proof: Let us use the notation of Usow (1967). Then $\underline{A}_s = \underline{\beta}^o$, $L(\underline{A}_s, x)$ is the "Lagrangian" form such that residuals of $\underline{A}_s$ vanish at m linearly independent rows, $\underline{b}^j$. In the "Lagrangian" form, $L(\underline{A}_s, x)$ is the j^{th} column of the inverse matrix of the m linearly independent rows in X, $1 \leq j \leq m$. Let

$$\rho = \max\left\{\frac{\partial}{\partial\theta} S(\theta)\Big|_{\theta=0-} , -\frac{\partial}{\partial\theta} S(\theta)\Big|_{\theta=0+}\right\}$$

where $S(\theta) = F(\underline{\beta}^o + \theta\underline{d}) = R(\underline{A}_s + \theta\underline{b}^j)$ for some $1 \le j \le m$. Since $S(\theta)$ in a convex, piecewise linear function of θ, $\rho \le 0$ implies that $S(\theta) \ge S(0)$ for all θ. That is, $R(\underline{A}_s + \theta\underline{b}^j) \ge R(\underline{A}_s)$, for all θ. Since this is true for all $\underline{b}^j$'s in all "Lagrangian" forms at $\underline{A}_s$, by Theorem 3.2.1 and Corollary 3.2.2, $\underline{A}_s$ is an ℓ_1 estimate, i.e., $\underline{\beta}^o$ is an ℓ_1 estimate. □

A modified algorithm based on the method of Usow (1967) is as follows.

Initialization stage: We assume that $[\phi_j(x_i)]$ is full-rank. Let us start with any m linearly independent rows in $[\phi_j(x_i)]$, say $U = \{u_1, u_2, \ldots, u_m\}$, solve the system of equations

$$\sum_{j=1}^{m} \phi_j(u_i)a_j^o = f(u_i) , \quad 1 \le i \le m$$

for $\underline{A}_o = (a_1^o, a_2^o, \ldots, a_m^o)^T$, and proceed to the iteration stage with $k = 0$, $r = 0$, and $\ell = 1$.

Iteration stage:

(i) Set $r = r+1$. If $r > m$ then set $r=1$.

(ii) If $R(\underline{A}_k \pm \varepsilon \underline{b}^r) \geq R(\underline{A}_k)$ for all $\varepsilon > 0$[1], then go to (iii); otherwise set $\underline{A}_{k+1} = \underline{A}_k - \hat{\theta}\underline{b}^r$,[2] $k = k+1$, $\ell = 1$, and go to (i).

(iii) If $\ell \geq m$ then go to (iv); otherwise, set $\ell = \ell+1$ and go to (i).

(iv) If no new "Lagrangian" form at $\underline{A}_k$, then terminate the iteration[3]; otherwise, use a new "Lagrangian" form at $\underline{A}_k$, set $\ell = 1$, and go to (i).

---

[1]The current "Lagrangian" form at $\underline{A}_k$ is recorded and $\underline{b}^r$ is computed accordingly.

[2]See the discussion on the Usow's Lemma 4.3.

[3]By Theorem 3.2.1 and Corollary 3.2.2, $\underline{A}_k$ is an ℓ_1 estimate.

4. METHODS FOR COMPUTING ℓ_p ESTIMATES IN THE LINEAR MODEL WHEN $p > 1$ AND $p \neq 2$

In this chapter, we will discuss some methods for computing an ℓ_p estimate in the linear model when $p > 1$ and $p \neq 2$. One way is to use Newton's method to solve the normal equations of the ℓ_p estimation problem. But the method may not converge when $p > 2$ and can have numerical difficulties when $1 < p < 2$. Hence, many people have proposed modifications of Newton's method and use of alternative methods. The quasi-Newton methods which require only the first order gradient are generally thought to be the best among available first order gradient methods for general minimization problems. They presumably sometimes work well for the ℓ_p estimation problem. We will propose a new method for the ℓ_p estimation problem. Since the method does not require a lot of computation, it is very fast. Next, we will derive a closed form solution for the ℓ_p estimation problem when the design matrix X is of dimension $(m + 1) \times m$. Under some conditions we can apply linear search and can derive a closed form solution for the ℓ_p estimation problem when X is of dimension $(m + 2) \times m$. Finally, we will discuss two different methods of generating test problems for the ℓ_p estimation problem.

An outline of the chapter is given as follows. We will first discuss Newton's method, a modified Newton's method, a method proposed by Ekblom (1973) using a modified Newton's method in a sequence of ℓ_p estimation problems formed by introducing various small perturbations

on the objective function, and a quasi-Newton method known as the Davidon-Fletcher-Powell method all in Section 4.1. The new method for the ℓ_p estimation problem will be discussed in Section 4.2. Closed form solutions for the ℓ_p estimation problems when the design matrix X is of dimension $(m + 1) \times m$ or $(m + 2) \times m$ will be given in Section 4.3. Two methods of generating test problems for the ℓ_p estimation problem will be given in Section 4.4.

To establish notation and set the stage for subsequent description, we now derive the gradient vector and the Hessian matrix of the objective function for the ℓ_p estimation problem. Let the objective function be denoted by

$$F(\underline{\beta}) = \sum_{i=1}^{n} |y_i - \underline{x}_i^T \underline{\beta}|^p$$

as in Chapter 1, where $p > 1$ and $p \neq 2$. As indicated in Section 2.1, we can assume a full-rank X matrix for our computing oriented discussion herein. Then there is a unique ℓ_p estimate for the full-rank ℓ_p estimation problem as discussed in Section 2.3. Let $\underline{g}(\underline{\beta})$ denote the gradient vector at $\underline{\beta}$, and $r_i = y_i - \underline{x}_i^T \underline{\beta}$, $1 \leq i \leq n$. Then, the gradient vector is expressible in terms of residuals as

$$\begin{aligned}
\underline{g}(\underline{\beta}) &= \frac{\partial}{\partial \underline{\beta}} F(\underline{\beta}) \\
&= - p \sum_{i=1}^{n} |y_i - \underline{x}_i^T \underline{\beta}|^{p-1} \operatorname{sign}(y_i - \underline{x}_i^T \underline{\beta}) \underline{x}_i \\
&= - p \sum_{i=1}^{n} |r_i|^{p-1} \operatorname{sign}(r_i) \underline{x}_i \\
&= - p X^T \underline{w}
\end{aligned}$$

where $\underline{w} = (w_1, w_2, \ldots, w_n)^T$ and $w_i = |r_i|^{p-1}\text{sign}(r_i)$, $1 \le i \le n$. Note that $\underline{g}(\underline{\beta})$ is well-defined for all $\underline{\beta} \in \mathbb{R}^m$.

Next let $H(\underline{\beta})$ denote the Hessian matrix at $\underline{\beta}$. Then $H(\underline{\beta})$ can be written as

$$H(\underline{\beta}) = \frac{\partial}{\partial \underline{\beta}} \underline{g}(\underline{\beta})$$

$$= p\ (p-1) \sum_{i=1}^{n} |r_i|^{p-2} \underline{x}_i\ \underline{x}_i^T$$

$$= p\ (p-1)\ X^T R\ X$$

where $R = \text{diag}(|r_i|^{p-2})$. Note that, when $1 < p < 2$, $H(\underline{\beta})$ is undefined if there is any zero residual at $\underline{\beta}$. Also, when $p > 2$, $H(\underline{\beta})$ is singular if there are more than $n-m$ zero residuals at $\underline{\beta}$.

4.1 Available Computational Methods for the ℓ_p Estimation Problem

We will discuss Newton's method and a modified Newton's method first. Newton's method is an iterative procedure such that in the k-th iteration, applied to our problem, we have

$$\underline{\beta}^{(k+1)} = \underline{\beta}^{(k)} - H^{(k)^{-1}} \underline{g}^{(k)}$$

where $H^{(k)} = H(\underline{\beta}^{(k)})$, and $\underline{g}^{(k)} = \underline{g}(\underline{\beta}^{(k)})$. Hence, in the k-th iteration,

$$H^{(k)}(\underline{\beta}^{(k+1)} - \underline{\beta}^{(k)}) = -\underline{g}^{(k)},$$

which becomes the following equations

$$p(p-1)X^T R^{(k)} X(\underline{\beta}^{(k+1)} - \underline{\beta}^{(k)}) = p\, X^T \underline{w}^{(k)}$$

where $R^{(k)}$ and $\underline{w}^{(k)}$ denote R and $\underline{w}$ at $\underline{\beta}^{(k)}$, respectively. Therefore, in the k-th iteration,

$$\underline{\beta}^{(k+1)} = \underline{\beta}^{(k)} + \frac{1}{p-1}\, \underline{d}^{(k)}$$

where $\underline{d}^{(k)}$ satisfies the equation

$$X^T R^{(k)} X\, \underline{d}^{(k)} = X^T \underline{w}^{(k)} .$$

By setting $\underline{g}(\underline{\beta}) = \underline{0}$, we have $X^T \underline{w} = \underline{0}$. Since $\underline{w} = R(\underline{y} - X\underline{\beta})$, we have

$$X^T R X \underline{\beta} = X^T R \underline{y}$$

which is referred to as the normal equations of the ℓ_p estimation problem. In the k-th iteration, the normal equations take the form

$$X^T R^{(k)} X \underline{\beta}^{(k+1)} = X^T R^{(k)} \underline{y} .$$

If we subtract $X^T R^{(k)} X$ from both sides of the above equations we get

$$X^T R^{(k)} X(\underline{\beta}^{(k+1)} - \underline{\beta}^{(k)}) = X^T \underline{w}^{(k)}$$

which is the same system of linear equations used to solve for $\underline{d}^{(k)}$ in the k-th iteration of Newton's method. Hence, we actually solve the normal equations of the ℓ_p estimation problem in each iteration of Newton's method.

It has been shown that Newton's method converges if the Hessian matrix remains positive-definite in each iteration. Barrodale and Roberts (1970) reported that Newton's method is about 10 times faster

than other convex programming methods they used, but could have numerical difficulties when $1 < p < 2$. As indicated earlier, when $1 < p < 2$, the Hessian matrix is undefined if there is any zero residual at the parameter vector. Also, when $p > 2$, the Hessian matrix is singular if there are more than $n-m$ zero residuals at the parameter vector. Kennedy and Gentle (1978), following the approach used by Merle and Spath (1974) in the iteratively reweighted least squares method for the ℓ_p estimation problem, changed the absolute value of zero residual or a residual which is close to zero to a preassigned lower bound in the Hessian matrix in each iteration, and tried to overcome the above difficulties. They found that this was not a very satisfactory procedure. The result from their work shows that the modified Newton's method works well in many cases for $p > 2$, but is not numerically stable for $1 < p < 2$.

Ekblom (1973) introduced a perturbation in $F(\underline{\beta})$ such that the objective function becomes $\sum_{i=1}^{n} ((y_i - \underline{x}_i^T\underline{\beta})^2 + e^2)^{\frac{p}{2}}$ and suggested using a modified Newton's method on a sequence of problems in which e^2 is decreased to zero. He adapted a Goldstein-Armijo steplength $\gamma^{(k)}$ and computed

$$\underline{\beta}^{(k+1)} = \underline{\beta}^{(k)} + \frac{\gamma^{(k)}}{p-1}\,\underline{d}^{(k)}$$

in the k-th iteration of the modified Newton's method. He found the method works well on many problems investigated. He concluded that the method can assure convergence to the optimum parameter vector for a problem very close to the original problem and can increase the rate of

convergence since the objective function of the problem with perturbation is smoother than the objective function of the original problem.

We now discuss a group of methods for the ℓ_p estimation problem which do not require the Hessian matrix. Hence, most of the difficulties encountered using Newton's method can be avoided. The quasi-Newton methods use an approximation to the inverse of the Hessian matrix which becomes closer to the true matrix as the iteration progresses. Fletcher and Powell (1963) have given a specific inverse update formula to generate approximation of the inverse of the Hessian matrix in each iteration of the method. They showed that the matrix generated by the inverse update formula in each iteration is symmetric and positive-definite if the matrix set in the first iteration is symmetric and positive-definite. Also, they proved that the method is quadratically convergent. (Details are discussed in Chapter 10 of Kennedy and Gentle (1980).) The inverse update formula forms the basis of the well-known Davidon-Fletcher-Powell method which is one of the most often used methods for general nonlinear minimization problems. Forsythe (1972) reported satisfactory results using the method for the ℓ_p estimation problem when $1 < p < 2$. Kennedy and Gentle (1978) studied methods for the ℓ_p estimation problem, $p > 1$ and $p \neq 2$, and found the method works well in general and is most desirable when $1 < p < 2$. Money, Affleck-Graves, Hart and Barr (1982) used the method in their Monte Carlo study for the choice of p in the ℓ_p estimation problem. However, the quasi-Newton method requires a large amount of space in computer memory for fair size

problems and requires linear search in each iteration. Both requirements are costly which makes the them undesirable in most applications.

4.2 The New Computing Method for the ℓ_p Estimation Problem

We will first describe a method for the ℓ_p estimation computing problem when the design matrix is of dimension $(m + 1) \times m$. The method also leads to a closed form solution for the ℓ_p estimation problem in this special case. Then, we will discuss the ℓ_p estimation problem in general. We can extend the special case method to the general ℓ_p estimation problem and will propose an new iterative computing procedure. An algorithm for the new method will also be provided and some numerical results will be presented.

Let us now consider the linear model such that X is of dimension $(m + 1) \times m$. Let $X = (\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_m)$, where $\underline{c}_i \in \mathbb{R}^{m+1}$, $1 \leq i \leq m$. Note that we assume X is full-rank, hence, the vectors $\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_m$ are linearly independent. Let $\underline{d} \in \mathbb{R}^{m+1}$ be such that $\underline{d} \neq \underline{0}$ and $\underline{c}_i^T \underline{d} = 0$ for $1 \leq i \leq m$, in other words, $X^T \underline{d} = \underline{0}$. Let $\underline{d} = (d_1, d_2, \ldots, d_{m+1})^T$ and $\underline{z} = (z_1, z_2, \ldots, z_{m+1})^T$, where $z_i = |d_i|^{\frac{1}{p-1}} \operatorname{sign}(d_i)$, $1 \leq i \leq m+1$. Note that $\underline{z}^T \underline{d} = \sum_{i=1}^{m+1} |d_i|^{\frac{p}{p-1}} > 0$ since $\underline{d} \neq \underline{0}$. Suppose $\underline{z} = X\underline{\tau}$ for some $\underline{\tau} \in \mathbb{R}^m$, then

$$\begin{aligned} \underline{z}^T \underline{d} &= \underline{\tau}^T X^T \underline{d} \\ &= \underline{\tau}^T \underline{0} \\ &= 0 \quad , \end{aligned}$$

which is a contradiction. Hence, $\underline{z} \notin C(X)$, the column space of X. Thus, $\{\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_m, \underline{z}\}$ is a set of $m+1$ linearly independent vectors in $\mathbb{R}^{m+1}$ and is a basis of $\mathbb{R}^{m+1}$. Therefore, there exists a $\underline{\beta} \in \mathbb{R}^m$ and an $\alpha \in \mathbb{R}$ such that $\underline{y} = X\underline{\beta} + \alpha\underline{z}$. Note that

$$\begin{aligned} w_i &= |r_i|^{p-1}\text{sign}(r_i) \\ &= |\alpha|^{p-1}\text{sign}(\alpha)|z_i|^{p-1}\text{sign}(z_i) \\ &= |\alpha|^{p-1}\text{sign}(\alpha)d_i \end{aligned}$$

for $1 \leq i \leq m+1$. Hence,

$$\begin{aligned} \underline{g}(\underline{\beta}) &= -\,p\,X^T\underline{w} \\ &= -\,p\,|\alpha|^{p-1}\text{sign}(\alpha)\,X^T\underline{d} \\ &= \underline{0}\,. \end{aligned}$$

Since the gradient vector at $\underline{\beta}$ is the zero vector, $\underline{\beta}$ is the ℓ_p estimate.

In general, we have $n > m+1$. We can find a $\underline{d} \in \mathbb{R}^n$ such that $X^T\underline{d} = \underline{0}$ and define $\underline{z} \in \mathbb{R}^n$ in the same way as above. We can prove $\underline{z} \notin C(X)$ similarly. Then, $\{\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_m, \underline{z}\}$ is a set of $m+1$ linearly independent vectors in $\mathbb{R}^n$ but is not a basis of $\mathbb{R}^n$. Hence, it is not necessary to have a $\underline{\beta} \in \mathbb{R}^m$ and an $\alpha \in \mathbb{R}$ such that $\underline{y} = X\underline{\beta} + \alpha\underline{z}$. If we use the least squares criterion such that $\underline{y} = X\hat{\underline{\beta}} + \hat{\alpha}\underline{z} + \hat{\underline{e}}$ and $\sum_{i=1}^{n} \hat{e}_i^2$ is minimized, where $\hat{\underline{e}} = (\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_n)^T$, we can get an approximation of the ℓ_p estimate, i.e., $\hat{\underline{\beta}}$. Now, let $\underline{r}(\varepsilon) = \hat{\alpha}\underline{z} + \varepsilon\,\hat{\underline{e}}$, where $0 < \varepsilon \leq 1$. Note that $\underline{r}(1) = \underline{y} - X\hat{\underline{\beta}}$, which

is the residual vector for the approximated estimate $\hat{\beta}$. Let $\underline{r}(\varepsilon) = (r_1(\varepsilon), r_2(\varepsilon), \ldots, r_n(\varepsilon))^T$ and $\underline{w}^o = (w_1^o, w_2^o, \ldots, w_n^o)^T$, where $w_i^o = |r_i(\varepsilon)|^{p-1}\text{sign}(r_i(\varepsilon))$, $1 \leq i \leq n$.

If $\varepsilon = 0$ were chosen, then

$$w_i^o = |\hat{\alpha}|^{p-1}\text{sign}(\hat{\alpha})|z_i|^{p-1}\text{sign}(z_i)$$

$$= |\hat{\alpha}|^{p-1}\text{sign}(\hat{\alpha})|d_i|\ \text{sign}(d_i)$$

$$= |\hat{\alpha}|^{p-1}\text{sign}(\hat{\alpha})d_i$$

for $1 \leq i \leq n$, in other words, $\underline{w}^o = |\hat{\alpha}|^{p-1}\text{sign}(\hat{\alpha})\underline{d}$. Hence, $\underline{w}^o$ is in the same direction as $\underline{d}$ which is the vector we start with. Note that the larger ε is, the closer to hyperplane $\{\underline{y} - X\underline{\beta} | \underline{\beta} \in \mathbb{R}^m\}$ $\underline{r}(\varepsilon)$ will be. Now, we would like to check to see whether $X^T\underline{w}^o = \underline{0}$. If $X^T\underline{w}^o$ is close to the zero vector then the corresponding $\underline{\beta}$ is computed for the ℓ_p estimate and the procedure is terminated. Otherwise, we need to adjust $\underline{w}^o$ toward a descent direction, in other words, change $\underline{w}^o$ to $\underline{w}$ such that $X^T\underline{w}$ is closer to the zero vector than is $X^T\underline{w}^o$.

Let us consider the fitting problem $X\underline{\delta} \cong \underline{w}^o$ and compute the least squares estimate for the problem. Thus, we solve $X^TX\hat{\underline{\delta}} = X^T\underline{w}^o$ for $\hat{\underline{\delta}}$. Let $\underline{w}(\varepsilon') = \underline{w}^o - \varepsilon' X\hat{\underline{\delta}}$, where $0 < \varepsilon' \leq 1$. Since

$$X^T\underline{w}(\varepsilon') = X^T(\underline{w}^o - \varepsilon' X\hat{\underline{\delta}})$$

$$= (1 - \varepsilon')X^T\underline{w}^o ,$$

where $0 \leq 1-\varepsilon' < 1$, $X^T\underline{w}(\varepsilon')$ is closer to the zero vector than is

$X^T\underline{w}^o$. Note that $X^T\underline{w}(1) = \underline{0}$. Hence, vector $\underline{w}(1)$ is a proper choice of the above $\underline{d}$. Also note that, the larger ε' is, the closer to the zero vector $X^T\underline{w}(\varepsilon')$ will be. We now have finished with one iteration of the procedure. Setting $\underline{d} = \underline{w}(\varepsilon')$ we then start the next iteration.

Numerical instability occurs if the residual is close to zero when $1 < p < 2$. We will discuss the problem and will suggest some changes in the method. Let $f(r)$ be a real-valued function defined as $f(r) = |r|^{p-1}\text{sign}(r)$. Note that $w_i^o = f(r_i)$, $1 \leq i \leq n$. Now $\frac{\partial}{\partial r} f(r) = (p-1)|r|^{p-2}$ for $r \neq 0$ when $1 < p < 2$. Since $\frac{\partial}{\partial r} f(r)$ exists for $r > 0$, by Taylor's Theorem,

$$f(r + \Delta r) = f(r) = \frac{\partial}{\partial \tilde{r}} f(r + \Delta\tilde{r})\Delta r ,$$

where $r > 0$, $r + \Delta r > 0$, Δr is close to zero, and $\Delta\tilde{r}$ lies between 0 and Δr . Note that $\frac{\partial}{\partial r} f(r + \Delta\tilde{r})$ may become large if r is close to zero when $1 < p < 2$. Hence, a slight change on r_i can cause a drastic change in w_i^o if r_i is positive and is close to zero when $1 < p < 2$. Similarly, a slight change on r_i can cause a drastic change in w_i^o if r_i is negative and is close to zero when $1 < p < 2$. Therefore, we set $r_i = \hat{\alpha}z_i$ if $\hat{\alpha}z_i$ is close to zero, $1 \leq i \leq n$, when $1 < p < 2$. Then, we will not have a drastic change in the direction of $\underline{w}^o$ from the direction of $\underline{d}$ when $1 < p < 2$. Furthermore, w_i of value close to zero would also make the method numerically unstable when $p > 2$. Let f^{-1} denote the inverse function of f , i.e., $f^{-1}(w) = |w|^{\frac{1}{p-1}}\text{sign}(w)$. Note that $0 < \frac{1}{p-1} < 1$ when $p > 2$. Also,

note that $z_i = f^{-1}(w_i)$, $1 \le i \le n$. Now, $\frac{\partial}{\partial w} f^{-1}(w) = \frac{1}{p-1} |w|^{\frac{1}{p-1} - 1}$ for $w \neq 0$ when $p > 2$. Since $\frac{\partial}{\partial w} f^{-1}(w)$ exists for $w > 0$, by Taylor's Theorem

$$f^{-1}(w + \Delta w) = f^{-1}(w) + \frac{\partial}{\partial w} f^{-1}(w + \Delta\tilde{w})\, \Delta w$$

where $w > 0$, $w + \Delta w > 0$, Δw is close to zero, and $\Delta\tilde{w}$ lies between 0 and Δw. Note that $\frac{\partial}{\partial w} f^{-1}(w + \Delta\tilde{w})$ may become large if w is close to zero when $p > 2$. Hence, a slight change on w_i can cause a drastic change in z_i if w_i is positive and is close to zero when $p > 2$. Similarly, a slight change on w_i can cause a drastic change in z_i if w_i is negative and is close to zero when $p > 2$. Therefore, we set $w_i = w_i^o$ if w_i^o is close to zero, $1 \le i \le n$, when $p > 2$. Then, we will not have a drastic change on $\underline{z}$ in the next iteration from $\underline{r}$ in the current iteration when $p > 2$.

We will propose an algorithm for the method discussed above as follows. (See Appendix for the FORTRAN code.)

Initialization stage: Let the fitting problem for ℓ_p estimation, $p > 1$ and $p \neq 2$, be $X\underline{\beta} \cong \underline{y}$ such that X is full-rank. We would like to start with the least squares estimate of $\underline{\beta}$. Hence, compute $\hat{\underline{\beta}} = (X^TX)^{-1}X^T\underline{y}$, $\hat{\underline{r}} = \underline{y} - X\hat{\underline{\beta}}$, and $\hat{\underline{w}}$, where $\hat{\underline{w}} = (\hat{w}_1, \hat{w}_2, \ldots, \hat{w}_n)^T$, $\hat{\underline{r}} = (\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_n)^T$, and $\hat{w}_i = |\hat{r}_i|^{p-1}\text{sign}(\hat{r}_i)$, $1 \le i \le n$. Set $S_1 = (X^TX)^{-1}$, $S_2 = (X^TX)^{-1}X^T$, $S_3 = I - X(X^TX)^{-1}X^T$, $S_4 = X^T\underline{y}$, and $\underline{w}^o = \hat{\underline{w}}$. Assign a small positive number to ξ and ξ', e.g., 10^{-13}, and proper values for ε and ε', then go to the following stage.

Iteration stage:

(i) Compute $\underline{u} = - X*S_2*\underline{w}^o$ (let $u_i = 0$ if $|w_i^o| \leq \xi$ when $p > 2$), $\underline{w} = \underline{w}^o + \varepsilon' \underline{u}$, and $\underline{z}$ (compute $z_i = |w_i|^{\frac{1}{p-1}} \text{sign}(w_i)$).

(ii) Sweep the last row of the matrix $\begin{pmatrix} S_1 & S_2*\underline{z} \\ -\underline{z}^T*S_2^T & \underline{z}^T*S_3*\underline{z} \end{pmatrix}$ to get the matrix V^{-1}[1], then compute $\begin{pmatrix} \underline{\beta} \\ \alpha \end{pmatrix} = V^{-1}\begin{pmatrix} S_4 \\ \underline{z}^T*\underline{y} \end{pmatrix}$, $\underline{e} = \underline{y} - (X\ \underline{z})\begin{pmatrix} \underline{\beta} \\ \alpha \end{pmatrix}$, (let $e_i = 0$ if $|\alpha z_i| \leq \xi$ when $1 < p < 2$) $\underline{r} = \alpha\underline{z} + \varepsilon\ \underline{e}$, and $\underline{w}^o$ (compute $w_i^o = |r_i|^{p-1}\text{sign}(r_i)$).

(iii) Compute $\bar{\underline{r}} = \underline{r} + (1 - \varepsilon)\ \underline{e}$, $\bar{\underline{w}}$ (compute $\bar{w}_i = |\bar{r}_i|^{p-1}\text{sign}(\bar{r}_i)$), and $\underline{g} = - p\ X^T*\bar{\underline{w}}$. If $\|\underline{g}\|_2 \leq \xi'$ then write $\underline{\beta}$ as the solution and terminate the procedure, otherwise go to (i).

We found the method converges with $\varepsilon = 0.2$ and $\varepsilon' = 1.0$ for $1.8 \leq p \leq 2.4$ on all problems investigated, and it converges with $\varepsilon = 0.4$ and $\varepsilon' = 1.0$ for $55 \geq p > 2$ and converges with $\varepsilon = 0.6$ and $\varepsilon' = 1.0$ for $1.4 \leq p < 2$ on most of problems investigated. Also, it converges with $\varepsilon = 1.0$ and $\varepsilon' = 0.4$ for $2 < p \leq 10$ and converges with $\varepsilon = 1.0$ and $\varepsilon' = 0.6$ for $1.6 \leq p < 2$ on some problems. Note that we can set ε or ε' a larger number to increase the speed of convergence. However, the method starts to diverge when ε or ε'

[1] In the fitting problem $(X\ \underline{z})\begin{pmatrix} \underline{\beta} \\ \alpha \end{pmatrix} \cong \underline{y}$, let $V = ((X\ \underline{z})^T(X\ \underline{z})) = \begin{pmatrix} X^TX & X^T\underline{z} \\ \underline{z}^TX & \underline{z}^T\underline{z} \end{pmatrix}$. We can obtain V^{-1} in the following way. We sweep the first m rows of V first to get the matrix $SV = \begin{pmatrix} (X^TX)^{-1} & (X^TX)^{-1}X^T\underline{z} \\ -\underline{z}^TX(X^TX)^{-1} & \underline{z}^T(I-X(X^TX)^{-1}X^T)\underline{z} \end{pmatrix}$, then sweep the last row of SV.

reaches some level. Also, we found the method became convergent after we standardized the design matrix X and the data vector $\underline{y}$ for some problems.

The proposed new method for the ℓ_p estimation problem is simple, can be easily coded for computer program, and converges rapidly for many cases. Hence, the method is desirable in application within the indicated range of p values.

4.3 A Method for Closed Form Solutions of the ℓ_p Estimation Problem with Special Cases of X

In this section, we will deal with the ℓ_p estimation problem when X is of dimension $(m + 1) \times m$ or $(m + 2) \times m$. By identifying a basis of $\mathbb{R}^m$ from rows in X and equating the gradient vector to the zero vector, we can derive a system of m linear equations in $\underline{\beta}$. In case X is of dimension $(m + 1) \times m$, we can show the system of equations is consistent. Hence, we can solve the system of equations for $\underline{\beta}$ and get a closed form solution. In case X is of dimension $(m+2) \times m$, under some conditions, we can apply linear search and form the system of equations which can be shown consistent. Hence, we also have essentially a closed form solution in this case.

We will start with the case such that X is of dimension $(m + 1) \times m$. Let $X = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_{m+1})^T$. Since X is full-rank, there exists a set of m rows in X, say $\{\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m\}$, which is a basis of $\mathbb{R}^m$. Hence, $\underline{x}_{m+1} = \sum_{i=1}^{m} a_i \underline{x}_i$. As indicated earlier,

$$\underline{g}(\underline{\beta}) = - p \sum_{i=1}^{m+1} |r_i|^{p-1} \text{sign}(r_i)\, \underline{x}_i$$

where $r_i = y_i - \underline{x}_i^T \underline{\beta}$, $1 \le i \le m+1$. By setting $\underline{g}(\underline{\beta}) = \underline{0}$ we have

$$\sum_{i=1}^{m+1} |r_i|^{p-1} \text{sign}(r_i)\, \underline{x}_i = \underline{0}$$

hence,

$$\sum_{i=1}^{m} (|r_i|^{p-1} \text{sign}(r_i) + a_i\, |r_{m+1}|^{p-1} \text{sign}(r_{m+1}))\, \underline{x}_i = \underline{0} \ .$$

Since $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m$ are linearly independent, we have a system of equations

$$|r_i|^{p-1} \text{sign}(r_i) + a_i\, |r_{m+1}|^{p-1} \text{sign}(r_{m+1}) = 0$$

for $1 \le i \le m$. In case $r_{m+1} = 0$, then $r_i = 0$ for $1 \le i \le m$. Hence, the ℓ_p estimate can be found by solving the system of equations $r_i = 0$, $1 \le i \le m$, which is the system of linear equations in $\underline{\beta}$

$$\underline{x}_i^T \underline{\beta} = y_i \ , \ 1 \le i \le m \ .$$

Note that it is consistent since $(\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m)^T$ is nonsingular. Also, the objective function has zero value at the solution in this case. In case $r_{m+1} \neq 0$, we divide the equations by r_{m+1} and get the following equations

$$\left|\frac{r_i}{r_{m+1}}\right|^{p-1} \text{sign}\left(\frac{r_i}{r_{m+1}}\right) + a_i = 0 \ , \ 1 \le i \le m \ .$$

After a proper transformation, we get

$$r_i \pm (\pm\, a_i)^{\frac{1}{p-1}}\, r_{m+1} = 0 \ , \ 1 \le i \le m$$

where the signs are chosen to make $\pm a_i$ positive for each $1 \leq i \leq m$. Hence, the ℓ_p estimate can be found by solving the system of linear equations in $\underline{\beta}$,

$$(\underline{x}_i + c_i \underline{x}_{m+1})^T \underline{\beta} = y_i + c_i y_{m+1} \ , \ 1 \leq i \leq m$$

where $c_i = \pm (\pm a_i)^{\frac{1}{p-1}}$ for each $1 \leq i \leq m$. Note that the equations are consistent since $\{\underline{x}_i + c_i \underline{x}_{m+1} \ , \ 1 \leq i \leq m\}$ is a set of m linearly independent vectors in $\mathbb{R}^m$. We now prove this result as follows. Let $\sum_{i=1}^{m} s_i(\underline{x}_i + c_i \underline{x}_{m+1}) = \underline{0}$ for some s_i, $1 \leq i \leq m$. Then,

$$\sum_{i=1}^{m} s_i \underline{x}_i = - \left(\sum_{i=1}^{m} s_i c_i \right) \underline{x}_{m+1}$$

$$= - \left(\sum_{j=1}^{m} s_j c_j \right) \left(\sum_{i=1}^{m} a_i \underline{x}_i \right)$$

$$= - \sum_{i=1}^{m} a_i \left(\sum_{j=1}^{m} s_j c_j \right) \underline{x}_i$$

hence,

$$s_i = - a_i \sum_{j=1}^{m} s_j c_j \ , \ 1 \leq i \leq m \ .$$

Multiplying both sides by c_i, $1 \leq i \leq m$, respectively and summing we get

$$\sum_{i=1}^{m} s_i c_i = - \left(\sum_{i=1}^{m} a_i c_i \right) \left(\sum_{j=1}^{m} s_j c_j \right) \ .$$

Suppose $\sum_{i=1}^{m} s_i c_i \neq 0$, then we can divide both sides by $\sum_{i=1}^{m} s_i c_i$ and get $\sum_{i=1}^{m} a_i c_i = -1$. However, $a_i c_i = \pm a_i (\pm a_i)^{\frac{1}{p-1}} \geq 0$, according

to the signs we have chosen to make $\pm a_i$ positive, $1 \le i \le m$. Hence, $\sum_{i=1}^{m} a_i c_i \ge 0$, which leads to a contradiction. Thus, $\sum_{i=1}^{m} s_i c_i = 0$, which implies $s_i = 0$, $1 \le i \le m$. Therefore, by definition, $\{\underline{x}_i + c_i \underline{x}_{m+1}, 1 \le i \le m\}$ is a set of m linearly independent vectors.

When X is of dimension $(m + 2) \times m$, similarly we assume $\{\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m\}$ is a basis of $\mathbb{R}^m$ and let $\underline{x}_{m+1} = \sum_{i=1}^{m} a_i \underline{x}_i$ and $\underline{x}_{m+2} = \sum_{i=1}^{m} b_i \underline{x}_i$. By setting $\underline{g}(\underline{\beta}) = \underline{0}$, similarly we can derive the system of equation

$$|r_i|^{p-1}\text{sign}(r_i) + a_i|r_{m+1}|^{p-1}\text{sign}(r_{m+1}) + b_i|r_{m+2}|^{p-1}\text{sign}(r_{m+2}) = 0,$$

$1 \le i \le m$. In case either $r_{m+1} = 0$ or $r_{m+2} = 0$, these equations are in the same forms as in the above discussion. Hence, we can solve the ℓ_p estimate accordingly. Otherwise, they take the following forms

$$\left|\frac{r_i}{r_{m+2}}\right|^{p-1} \text{sign}\left(\frac{r_i}{r_{m+2}}\right) + a_i \left|\frac{r_{m+1}}{r_{m+2}}\right|^{p-1} \text{sign}\left(\frac{r_{m+1}}{r_{m+2}}\right) + b_i = 0$$

$1 \le i \le m$, which is the same as the system of equations

$$\lambda_i + a_i \lambda_{m+1} + b_i = 0 , \quad 1 \le i \le m$$

and

$$\lambda_i = \left|\frac{r_i}{r_{m+2}}\right|^{p-1} \text{sign}\left(\frac{r_i}{r_{m+2}}\right) , \quad 1 \le i \le m+1 .$$

After a proper transformation, we have the following system of equations

$$\lambda_i + a_i \lambda_{m+1} + b_i = 0 , \quad 1 \le i \le m$$

and

$$r_i \mp (\pm \lambda_i)^{\frac{1}{p-1}} r_{m+2} = 0 , \quad 1 \le i \le m+1$$

where the signs are chosen to make $\pm \lambda_i$ positive, $1 \leq i \leq m+1$. Note that once λ_{m+1} is known, the ℓ_p estimate can be obtained by solving the system of m equations

$$r_i \mp (\pm \lambda_i)^{\frac{1}{p-1}} r_{m+2} = 0 \ , \ 1 \leq i \leq m$$

which is the system of m linear equations in $\underline{\beta}$

$$(\underline{x}_i + d_i \underline{x}_{m+2})^T \underline{\beta} = y_i + d_i y_{m+2} \ , \ 1 \leq i \leq m \ ,$$

where $d_i = \mp (\pm \lambda_i)^{\frac{1}{p-1}}$, $1 \leq i \leq m$. Note that it is consistent when $\{\underline{x}_i + d_i \underline{x}_{m+2} \ , \ 1 \leq i \leq m\}$ is a set of m linearly independent vectors in $\mathbb{R}^m$. We will prove that $\{\underline{x}_i + d_i \underline{x}_{m+2} \ , \ 1 \leq i \leq m\}$ is a set of m linearly independent vectors if λ_{m+1} is close to zero as follows. Let $\sum_{i=1}^{m} s_i(\underline{x}_i + d_i \underline{x}_{m+2}) = \underline{0}$ for some s_i , $1 \leq i \leq m$. Then,

$$\sum_{i=1}^{m} s_i \underline{x}_i = - \left(\sum_{i=1}^{m} s_i d_i \right) \underline{x}_{m+2}$$

$$= - \left(\sum_{j=1}^{m} s_j d_j \right) \left(\sum_{i=1}^{m} b_i \underline{x}_i \right)$$

$$= - \sum_{i=1}^{m} b_i \left(\sum_{j=1}^{m} s_j d_j \right) \underline{x}_i$$

hence,

$$s_i = - b_i \sum_{j=1}^{m} s_j d_j \ , \ 1 \leq i \leq m \ .$$

Multiplying both sides by d_i , $1 \leq i \leq m$, respectively and summing, we get

$$\sum_{i=1}^{m} s_i d_i = - \left(\sum_{i=1}^{m} b_i d_i \right) \left(\sum_{j=1}^{m} s_j d_j \right) \ .$$

Suppose $\sum_{i=1}^{m} s_i d_i \neq 0$, then we can divide both sides by $\sum_{i=1}^{m} s_i d_i$ and get $\sum_{i=1}^{m} b_i d_i = -1$. However,

$$b_i d_i = -(\lambda_i + a_i \lambda_{m+1}) d_i$$

$$= \pm (\lambda_i + a_i \lambda_{m+1})(\pm \lambda_i)^{\frac{1}{p-1}}$$

$$\geq 0$$

when λ_{m+1} is close to zero, for $1 \leq i \leq m$. Hence, $\sum_{i=1}^{m} b_i d_i \geq 0$, which leads to a contradiction. Thus, $\sum_{i=1}^{m} s_i d_i = 0$, which implies $s_i = 0$, $1 \leq i \leq m$. Therefore, by definition, $\{\underline{x}_i + d_i \underline{x}_{m+2}$, $1 \leq i \leq m\}$ is a set of m linearly independent vectors.

We now derive an equation containing only the unknown variable λ_{m+1}. Then we can apply linear search to find λ_{m+1} satisfying the equation. We would assume that λ_{m+1} is close to zero. Note that the equation

$$r_{m+1} \mp (\pm \lambda_{m+1})^{\frac{1}{p-1}} r_{m+2} = 0$$

is the same as the following equation

$$(\underline{x}_{m+1} + d_{m+1} \underline{x}_{m+2})^T \underline{\beta} = y_{m+1} + d_{m+1} y_{m+2}$$

where $d_{m+1} = \mp (\pm \lambda_{m+1})^{\frac{1}{p-1}}$. Since $\{\underline{x}_i + d_i \underline{x}_{m+2}$, $1 \leq i \leq m\}$ is a basis of $\mathbb{R}^m$ as shown earlier, there exist v_i, $1 \leq i \leq m$, such that

$$\sum_{i=1}^{m} v_i (\underline{x}_i + d_i \underline{x}_{m+2}) = \underline{x}_{m+1} + d_{m+1} \underline{x}_{m+2}$$

hence,

$$\sum_{i=1}^{m} (v_i + b_i \sum_{j=1}^{m} v_j d_j) \underline{x}_i = \sum_{i=1}^{m} (a_i + d_{m+1} b_i) \underline{x}_i .$$

Since the vectors $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m$ are linearly independent, we have

$$v_i + b_i \sum_{j=1}^{m} v_j d_j = a_i + d_{m+1} b_i , \quad 1 \leq i \leq m .$$

We can divide both sides by b_i , if $b_i \neq 0$, to get

$$\frac{v_i}{b_i} + \sum_{j=1}^{m} d_j v_j = \frac{a_i}{b_i} + d_{m+1}$$

and write $v_i = a_i$ if $b_i = 0$, for $1 \leq i \leq m$. Now, since the solution satisfies all $m+1$ equations

$$(\underline{x}_i + d_i \underline{x}_{m+2})^T \underline{\beta} = y_i + d_i y_{m+2}$$

we have

$$\sum_{i=1}^{m} v_i (y_i + d_i y_{m+2}) = y_{m+1} + d_{m+1} y_{m+2} .$$

We can divide both sides by y_{m+2} , if $y_{m+2} \neq 0$, to get

$$\sum_{i=1}^{m} (\frac{y_i}{y_{m+2}} + d_i) v_i = \frac{y_{m+1}}{y_{m+2}} + d_{m+1}$$

and write $\sum_{i=1}^{m} y_i v_i = y_{m+1}$ if $y_{m+2} = 0$. Since $\underline{x}_{m+2} \neq 0$, there exists at least one $b_i \neq 0$, $1 \leq i \leq m$, say $b_1 \neq 0$. Then, after proper subtractions, we have the following system of linear equation in $\underline{v}$, where $\underline{v}^T = (v_1, v_2, \ldots, v_m)$,

$$\frac{v_1}{b_1} - \frac{v_i}{b_i} = \frac{a_1}{b_1} - \frac{a_i}{b_i} , \quad (v_i = a_i , \text{ if } b_i = 0) , \quad 2 \leq i \leq m$$

$$\frac{v_1}{b_1} - \sum_{i=1}^{m} \frac{y_i}{y_{m+2}} v_i = \frac{a_1}{b_1} - \frac{y_{m+1}}{y_{m+2}} , \quad \left(\sum_{i=1}^{m} y_i v_i = y_{m+1} \right.$$

$$\left. \text{if } y_{m+2} = 0 \right).$$

Note that it is consistent, hence, we can solve for $\underline{v}$. Let us now consider the equation

$$\frac{v_1}{b_1} + \sum_{i=1}^{m} d_i v_i = \frac{a_1}{b_1} + d_{m+1}$$

where $d_i = \mp (\pm \lambda_i)^{\frac{1}{p-1}}$, $1 \leq i \leq m+1$, and $\lambda_i = - a_i \lambda_{m+1} - b_i$, $1 \leq i \leq m$. Since λ_i $(1 \leq i \leq m)$ can be expressed by a function of λ_{m+1} , the equation involves only the unknown variable λ_{m+1} . We then proceed to apply linear search to find the value for λ_{m+1} .

4.4 Two Methods of Generating Test Problems for ℓ_p Estimation in the Linear Model

Given any selected vector $\underline{\beta}^*$ in $\mathbb{R}^m$ and a full-rank matrix X , we can find a vector $\underline{y}$ in $\mathbb{R}^n$ such that $\underline{\beta}^*$ is the ℓ_p estimate of the parameter vector in the linear model $\underline{y} = X\underline{\beta} + \underline{e}$, where $p > 1$. This provides test problems for the ℓ_p estimation problem. We will describe two different methods of generating test problems and will verify the given $\underline{\beta}^*$ is indeed the ℓ_p estimate for the test problem generated.

We now describe one of the two methods as follows.

(i) Select a matrix X of dimension $n \times m$ such that $n > m$ and rank $(X) = m$.

(ii) Decompose X by the Gram-Schmidt orthogonalization method

such that $X = QT$, where Q and T are matrices of dimension $n \times m$ and $m \times m$ respectively, $Q^TQ = I_m$, and T is nonsingular and is in an upper triangular form.

(iii) Choose $\underline{a} \in \mathbb{R}^n$ arbitrarily and compute $\underline{w}^o$ as

$$\underline{w}^o = (I - QQ^T)\underline{a} .$$

(iv) Transform w_i^o as $r_i^o = |w_i^o|^{\frac{1}{p-1}} \text{sign}(w_i^o)$ and form the vector $\underline{r}^o$ with the r_i^o values.

(v) Select $\underline{\beta}^* \in \mathbb{R}^m$ and compute $\underline{y}$ as

$$\underline{y} = X\underline{\beta}^* + \underline{r}^o$$

We now verify that $\underline{\beta}^*$ is the ℓ_p estimate for the fitting problem $X\underline{\beta} \cong \underline{y}$. Let us consider the gradient vector at $\underline{\beta}^*$. The gradient vector $\underline{g}(\underline{\beta})$ takes the form $\underline{g}(\underline{\beta}) = -\, p\, X^T\underline{w}$, where $w_i = |r_i|^{p-1}\text{sign}(r_i)$ and $r_i = y_i - \underline{x}_i^T\underline{\beta}$, $1 \le i \le n$. Then, at $\underline{\beta} = \underline{\beta}^*$, $r_i^* = y_i - \underline{x}_i^T\underline{\beta}^* = r_i^o$ and $w_i^* = |r_i^*|^{p-1}\text{sign}(r_i^*) = |r_i^o|^{p-1}\text{sign}(r_i^o) = (|w_i^o|^{\frac{1}{p-1}})^{p-1}\text{sign}(w_i^o) = |w_i^o|\text{sign}(w_i^o) = w_i^o$, $1 \le i \le n$, in other words, $\underline{w}^* = \underline{w}^o$. Hence,

$$\begin{aligned} g(\underline{\beta}^*) &= -\, p\, X^T\underline{w}^* \\ &= -\, p\, X^T\underline{w}^o \\ &= -\, p\, X^T(I - QQ^T)\underline{a} \\ &= -\, p\, T^TQ^T(I - QQ^T)\underline{a} \\ &= -\, p\, T^T(Q^T - Q^T)\underline{a} \\ &= \underline{0} . \end{aligned}$$

Since the gradient vector at $\underline{\beta}^*$ is the zero vector , $\underline{\beta}^*$ is the ℓ_p estimate for the fitting problem $X\underline{\beta} \cong \underline{y}$.

Another method of generating test problems for the ℓ_p estimation problem is described as follows. Let $\underline{\beta}^* \in \mathbb{R}^m$ be given. Let $\{\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m\}$ be a basis of $\mathbb{R}^m$. Let $n > m$ and

$$\underline{x}_j = \sum_{i=1}^{m} a_{ji}\, \underline{x}_i \ , \quad m+1 \le j \le n \ .$$

Let $X = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m\ \underline{x}_{m+1} \cdots \underline{x}_n)^T$. Then X is a matrix of dimension $n \times m$ and is full-rank. We can choose y_j , $m+1 \le j \le n$ arbitrarily and compute $r_j = y_j - \underline{x}_j^T\underline{\beta}^*$, $m+1 \le j \le n$. Then, solve

$$|r_i|^{p-1}\mathrm{sign}(r_i) + \sum_{j=m+1}^{n} |r_j|^{p-1}\mathrm{sign}(r_j)a_{ji} = 0$$

for r_i and compute $y_i = r_i + \underline{x}_i^T\underline{\beta}^*$, $1 \le i \le m$. We now prove the given $\underline{\beta}^*$ is the ℓ_p estimate for the fitting problem $X\underline{\beta} \cong \underline{y}$. We would derive the gradient vector as follows.

$$\begin{aligned}
\underline{g}(\underline{\beta}) &= -\,p \sum_{i=1}^{n} |r_i|^{p-1}\mathrm{sign}(r_i)\underline{x}_i \\
&= -\,p\left(\sum_{i=1}^{m} |r_i|^{p-1}\mathrm{sign}(r_i)\underline{x}_i \right. \\
&\qquad \left. + \sum_{i=m+1}^{n} |r_i|^{p-1}\mathrm{sign}(r_i) \sum_{j=1}^{m} a_{ij}\, \underline{x}_j \right) \\
&= -p\left(\sum_{i=1}^{m} |r_i|^{p-1}\mathrm{sign}(r_i)\underline{x}_i \right. \\
&\qquad \left. + \sum_{j=1}^{m} \sum_{i=m+1}^{n} |r_i|^{p-1}\mathrm{sign}(r_i)a_{ij}\, \underline{x}_j \right)
\end{aligned}$$

$$= - p \sum_{i=1}^{m} \left(|r_i|^{p-1} \operatorname{sign}(r_i) + \sum_{j=m+1}^{n} |r_j|^{p-1} \operatorname{sign}(r_j) a_{ji} \right) \underline{x}_i \, .$$

Since, at $\underline{\beta} = \underline{\beta}^*$, $|r_i|^{p-1} \operatorname{sign}(r_i) + \sum_{j=m+1}^{n} |r_j|^{p-1} \operatorname{sign}(r_j) a_{ji} = 0$,

$1 \leq i \leq m$, we have $\underline{g}(\underline{\beta}^*) = \underline{0}$. Hence, $\underline{\beta}^*$ is the ℓ_p estimate.

5. AUGMENTED LINEAR MODELS

The least squares estimate in the linear model with nonorthogonal data may be greatly improved, in the sense of "total variation", by augmenting the design matrix X with a positive number multiple of the identity matrix I_m. Hoerl and Kennard (1970) showed that the total mean squared error is reduced significantly by the employment of the biased estimate. They called the method "Ridge Regression". Marquardt (1970) summarized some results on Ridge Regression and emphasized that the ridge estimate overcomes a serious deficiency of the least squares estimate by reducing the Euclidean length of the estimate in the parameter space.

Banks and Taylor (1980) used the least absolute deviation criterion in the augmented linear models to fit seismic processing data. They reported excellent results in goodness of fit. Hence, we are motivated to extend this work and investigate the criterion of the least ℓ_p norm of residual vector in the augmented linear model for any $p \geq 1$. Some results in ℓ_p estimation with the augmented linear model for $p > 1$ are given in this chapter. Among these we find that when the ℓ_p estimate is not unique, we can identify one ℓ_p estimate by considering a sequence of ℓ_p estimates for full-rank augmented linear models for $p > 1$, and with some additional conditions, for $p = 1$. We will discuss cases of $p = 2$ and $p = 1$ first in the following sections.

5.1 Introduction and Some Basic Properties

Let us first consider the linear model in Chapter 1 ,

$$\underline{y} = X\underline{\beta} + \underline{e} \quad .$$

Let

$$F(\underline{\beta}) = \sum_{i=1}^{n} |y_i - \underline{x}_i^T\underline{\beta}|^p \quad , \; p \geq 1$$

as in Chapter 1. We augment with the matrix $(\lambda^{\frac{1}{p}} I_m \; \underline{0})$, $\lambda > 0$, to obtain the problem

$$\begin{pmatrix} X \\ \hline \lambda^{\frac{1}{p}} I_m \end{pmatrix} \underline{\beta} \cong \begin{pmatrix} \underline{y} \\ \underline{0} \end{pmatrix} .$$

Then, the objective function for ℓ_p estimation in this problem is

$$\sum_{i=1}^{n} |y_i - \underline{x}_i^T\underline{\beta}|^p + \lambda \sum_{j=1}^{m} |\beta_j|^p \quad .$$

Note that we have taken into account the p^{th} power of the ℓ_p norm of the parameter vector in the minimization problem. If we put in weights $p_i > 0$, $1 \leq i \leq n$, $q_j > 0$, $1 \leq j \leq m$, to form the objective function as

$$\sum_{i=1}^{n} p_i |y_i - \underline{x}_i^T\underline{\beta}|^p + \lambda \sum_{j=1}^{m} q_j |\beta_j|^p$$

and by changing variables write $p_i^{\frac{1}{p}} y_i = y_i'$, $q_j^{\frac{1}{p}} \beta_j' = \beta_j'$, $\dfrac{p_i^{\frac{1}{p}} x_{ij}}{q_j^{\frac{1}{p}}} = x_{ij}'$

we have the objective function as

$$\sum_{i=1}^{n} |y_i' - \underline{x}_i'^T \underline{\beta}'|^p + \lambda \sum_{j=1}^{m} |\beta_j'|^p$$

which is in the same form as the previous objective function. Hence, we can let $p_i = 1$, $1 \leq i \leq n$, and $q_j = 1$, $1 \leq j \leq m$, without loss of generality. Note that the minimization problem is the Ridge Regression problem when $p = 2$ and is the problem Banks and Taylor (1980) considered when $p = 1$.

Use the notation

$$f(\underline{\beta}) = \sum_{j=1}^{m} |\beta_j|^p$$

the objective function is $F(\underline{\beta}) + \lambda f(\underline{\beta})$, where $\lambda > 0$. Let $\underline{\beta}^* \in \mathbb{R}^m$ such that $F(\underline{\beta}^*) \leq F(\underline{\beta})$ for all $\underline{\beta} \in \mathbb{R}^m$, and let S^* denote the set of all such $\underline{\beta}^*$. Also, let $\underline{\beta}_\lambda \in \mathbb{R}^m$ be such that $F(\underline{\beta}_\lambda) + \lambda f(\underline{\beta}_\lambda) \leq F(\underline{\beta}) + \lambda f(\underline{\beta})$ for all $\underline{\beta} \in \mathbb{R}^m$, and S_λ denote the set of all such $\underline{\beta}_\lambda$. By Lemma 2.3.2, it follows that $S^* \neq \emptyset$ and $S_\lambda \neq \emptyset$. Let $M = F(\underline{\beta}^*)$, in other words, $M = \inf\{F(\underline{\beta}) \mid \underline{\beta} \in \mathbb{R}^m\}$; and similarly define $L = \inf\{f(\underline{\beta}^*) \mid \underline{\beta}^* \in S^*\}$. We will use the same notation in the following sections unless otherwise specified. Now we introduce some basic properties in ℓ_p estimation with the augmented linear model.

Lemma 5.1.1. $f(\underline{\beta}_\lambda) \leq f(\underline{\beta}^*)$, hence, $f(\underline{\beta}_\lambda) \leq L$.

Proof: By definition,

$$F(\underline{\beta}^*) \leq F(\underline{\beta}_\lambda)$$

and $$F(\underline{\beta}_\lambda) + \lambda f(\underline{\beta}_\lambda) \leq F(\underline{\beta}^*) + \lambda f(\underline{\beta}^*) .$$

It follows that

$$F(\underline{\beta}^*) + F(\underline{\beta}_\lambda) + \lambda f(\underline{\beta}_\lambda) \leq F(\underline{\beta}_\lambda) + F(\underline{\beta}^*) + \lambda f(\underline{\beta}^*)$$

and $\lambda f(\underline{\beta}_\lambda) \leq \lambda f(\underline{\beta}^*)$.

Hence, $f(\underline{\beta}_\lambda) \leq f(\underline{\beta}^*)$, since $\lambda > 0$. □

Lemma 5.1.2. $F(\underline{\beta}_\lambda) \geq M$. The proof of this lemma follows from the definition of M .

Lemma 5.1.3. $f(\underline{\beta}_{\lambda_1}) \leq f(\underline{\beta}_{\lambda_2})$ when $\lambda_1 > \lambda_2$.

Proof: By definition,

$$F(\underline{\beta}_{\lambda_1}) + \lambda_1 f(\underline{\beta}_{\lambda_1}) \leq F(\underline{\beta}_{\lambda_2}) + \lambda_1 f(\underline{\beta}_{\lambda_2})$$

and $F(\underline{\beta}_{\lambda_2}) + \lambda_2 f(\underline{\beta}_{\lambda_2}) \leq F(\underline{\beta}_{\lambda_1}) + \lambda_2 f(\underline{\beta}_{\lambda_1})$.

Thus, it follows that

$$(\lambda_1 - \lambda_2) f(\underline{\beta}_{\lambda_1}) \leq (\lambda_1 - \lambda_2) f(\underline{\beta}_{\lambda_2}) .$$

Hence, $f(\underline{\beta}_{\lambda_1}) \leq f(\underline{\beta}_{\lambda_2})$, since $\lambda_1 > \lambda_2$. □

Lemma 5.1.4. $F(\underline{\beta}_{\lambda_1}) \geq F(\underline{\beta}_{\lambda_2})$ when $\lambda_1 > \lambda_2$.

Proof: From the definitions we have that

$$\frac{1}{\lambda_1} F(\underline{\beta}_{\lambda_1}) + f(\underline{\beta}_{\lambda_1}) \leq \frac{1}{\lambda_1} F(\underline{\beta}_{\lambda_2}) + f(\underline{\beta}_{\lambda_2})$$

and $$\frac{1}{\lambda_2} F(\underline{\beta}_{\lambda_2}) + f(\underline{\beta}_{\lambda_2}) \leq \frac{1}{\lambda_2} F(\underline{\beta}_{\lambda_1}) + f(\underline{\beta}_{\lambda_1}) .$$

Therefore, we can write

$$(\frac{1}{\lambda_2} - \frac{1}{\lambda_1})F(\underline{\beta}_{\lambda_1}) \geq (\frac{1}{\lambda_2} - \frac{1}{\lambda_1})F(\underline{\beta}_{\lambda_2}) .$$

Hence, $F(\underline{\beta}_{\lambda_1}) \geq F(\underline{\beta}_{\lambda_2})$ since $\frac{1}{\lambda_2} > \frac{1}{\lambda_1}$ □

5.2 Results from Ridge Regression

Some results that are well-known in Ridge Regression can be extended to ℓ_p estimation with the augmented linear model for $p > 1$ and $p \neq 2$. In this section, we state known results which will be extended in subsequent sections.

Similar results in Ridge Regression are given by many authors. The following two theorems are due to Marquardt (1970).

Theorem 5.2.1. $F(\underline{\beta}_\lambda) \leq F(\underline{\beta})$ for all $\underline{\beta} \in \mathbb{R}^m$ such that $f(\underline{\beta})^{\frac{1}{2}} \leq f(\underline{\beta}_\lambda)^{\frac{1}{2}}$. Furthermore, $F(\underline{\beta}_\lambda)$ is a monotone increasing function of λ (p=2 case.)

Theorem 5.2.2. $f(\underline{\beta}_\lambda)^{\frac{1}{2}}$ is a continuous monotone decreasing function of λ such that $f(\underline{\beta}_\lambda)^{\frac{1}{2}} \to 0$ as $\lambda \to \infty$ (p=2 case.)

When X is not full-rank there is not a unique least squares estimate in the linear model. However, there is only one least squares estimate having the least Euclidean length, namely, $X^+\underline{y}$, where X^+ is the Moore-Penrose inverse of X. Let

$$X = Q\begin{bmatrix} R_1 & 0 \\ 0 & 0 \end{bmatrix} U^T ,$$

as in Section 2.1. Then,

$$X^{+} = U\begin{pmatrix} R_1^{-1} & 0 \\ 0 & 0 \end{pmatrix} Q^{T} .$$

Thus,
$$X^{+}\underline{y} = U\begin{pmatrix} R_1^{-1} & 0 \\ 0 & 0 \end{pmatrix} Q^{T}\underline{y}$$

$$= (U_1U_2)\begin{pmatrix} R_1^{-1} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} Q_1^{T} \\ Q_2^{T} \end{pmatrix}\underline{y}$$

$$= U_1R_1^{-1}Q_1^{T}\underline{y} ,$$

where $U = (U_1U_2)$ and $Q = (Q_1Q_2)$. As indicated in Section 2.1, $\underline{\beta}^{*} = U\begin{pmatrix} \underline{a}_1^{*} \\ \underline{a}_2 \end{pmatrix}$ is a least squares estimate, where $\underline{a}_1^{*} = (S_1^{T}S_1)^{-1}S_1^{T}\underline{y}$ and $\underline{a}_2$ is arbitrary. Note that $S_1^{T}S_1 = R_1^{T}R_1$ and $S_1 = Q_1R_1$. Hence, $\underline{a}_1^{*} = R_1^{-1}Q_1^{T}\underline{y}$. Now,

$$\underline{\beta}^{*T}\underline{\beta}^{*} = (\underline{a}_1^{*T}\underline{a}_2^{T})U^{T}U\begin{pmatrix} \underline{a}_1^{*} \\ \underline{a}_2 \end{pmatrix}$$

$$= \underline{a}_1^{*T}\underline{a}_1^{*} + \underline{a}_2^{T}\underline{a}_2 .$$

Hence, $\underline{\beta}^{**} = U\begin{pmatrix} \underline{a}_1^{*} \\ \underline{0} \end{pmatrix}$ yields the least Euclidean length.

Since $\underline{\beta}^{**} = (U_1 U_2)\begin{pmatrix} \underline{a}_1^* \\ \underline{0} \end{pmatrix}$

$= U_1 \underline{a}_1^*$

$= U_1 R_1^{-1} Q_1^T \underline{y}$

$= X^+ \underline{y}$

$X^+ \underline{y}$ is the least squares estimate having the least Euclidean length.

Now, since $\begin{pmatrix} X \\ \sqrt{\lambda}\ I \end{pmatrix}$ is full-rank, we can solve the normal equations in the augmented linear model for $\underline{\beta}_\lambda$ to obtain

$$\underline{\beta}_\lambda = \left[\begin{pmatrix} X \\ \sqrt{\lambda}\ I \end{pmatrix}^T \begin{pmatrix} X \\ \sqrt{\lambda}\ I \end{pmatrix}\right]^{-1} \begin{pmatrix} X \\ \sqrt{\lambda}\ I \end{pmatrix} \begin{pmatrix} \underline{y} \\ \underline{0} \end{pmatrix}$$

$$= (X^T X + \lambda I)^{-1} X^T \underline{y} \ , \quad \text{where} \quad \lambda > 0 \ .$$

It can be shown that

$$\lim_{\sqrt{\lambda} \to 0} (X^T X + \lambda I)^{-1} X^T = X^+ \ .$$

Since $\lim_{\lambda \to 0} (X^T X + \lambda I)^{-1} X^T = \lim_{\sqrt{\lambda} \to 0} (X^T X + \lambda I)^{-1} X^T$, where $\lambda > 0$,

we have $\lim_{\lambda \to 0} \underline{\beta}_\lambda = \lim_{\lambda \to 0} (X^T X + \lambda I)^{-1} X^T \underline{y} = X^+ \underline{y}$. Thus, the sequence of least squares estimates for augmented linear models converges to the least squares estimate having the least Euclidean norm for the original linear model as λ approaches zero. This result can be extended to the cases of $p > 1$ and $p \neq 2$, and also with some conditions to the case of $p = 1$.

5.3 Properties of Limiting $\underline{\beta}_\lambda$ as λ Approaches Zero for the Case $p = 1$

Banks and Taylor (1980) applied the least absolute deviation technique with the augmented linear model to geophysical problems involving seismic processing. They modified the simplex method of Barrodale and Roberts (1973) so that the computer memory requirement of the method is reduced in order to make it more efficient to handle large scale seismic processing problem. They are able to recover the original seismic spike train from the modeled seismic trace. They reported satisfactory results when λ was chosen close to zero. They also pointed out the importance of including terms with λ in the objective function since they failed to recover any spike train if they use the original linear model. In our notation, the result they found is that $\underline{\beta}_\lambda$, with sufficiently small λ, is much more desirable than $\underline{\beta}^*$ in seismic processing problems. Note that, as indicated in Section 2.3, both $\underline{\beta}_\lambda$ and $\underline{\beta}^*$ are not necessarily unique. However, under some conditions $\underline{\beta}_\lambda$ is unique and the sequence of $\underline{\beta}_\lambda$ converges to a uniquely determined $\underline{\beta}^*$ as λ approaches zero. We proceed to discuss what these conditions are and the form of the uniquely determined $\underline{\beta}^*$.

S^* is closed, since if $\tilde{\underline{\beta}}$ is a cluster vector of S^*, then there is a sequence of vectors in S^*, say $\{\underline{\beta}_i^*\}_{i=1}^{\infty}$, such that $\lim_{i\to\infty} \underline{\beta}_i^* = \tilde{\underline{\beta}}$, i.e., $\lim_{i\to\infty} M = F(\tilde{\underline{\beta}})$, which implies $F(\tilde{\underline{\beta}}) = M$, i.e., $\tilde{\underline{\beta}} \in S^*$. Thus, $S^{**} = \{\underline{\beta}^{**} \in S^* | f(\underline{\beta}^{**}) \leq f(\underline{\beta}^*)$ for all $\underline{\beta}^* \in S^*\} \neq \emptyset$. Note that $\beta^{**} \in S^*$ and $f(\underline{\beta}^{**}) = L$ if and only if $\underline{\beta}^{**} \in S^{**}$.

Theorem 5.3.1. If $\underline{\beta}^{**} \in S^{**}$ is such that residuals of $\underline{\beta}^{**}$ vanish at m linearly independent rows in the augmented design matrix, then $\underline{\beta}^{**} \in S_\lambda$ when λ is sufficiently small.

Proof: Let Z be the inverse of matrix of any m linearly independent rows at which residuals of $\underline{\beta}^{**}$ vanish. Let $\underline{d}$ be any column in Z. Let

$$S(\theta) = F(\underline{\beta}^{**} + \theta\underline{d}) + \lambda f(\underline{\beta}^{**} + \theta\underline{d}) .$$

Note that $S(\theta)$ is a convex, piecewise linear function of θ. Let $\rho = \max(\frac{\partial}{\partial\theta} S(\theta)|_{\theta=0-}, -\frac{\partial}{\partial\theta} S(\theta)|_{\theta=0+})$. If $\rho \leq 0$ for all such $\underline{d}$, then $\underline{\beta}^{**} \in S_\lambda$, by Corollary 3.2.3. Hence, it is sufficient to prove that $\rho \leq 0$ for all such $\underline{d}$.

$$\frac{\partial}{\partial\theta} S(\theta)|_{\theta=0-} = \frac{\partial}{\partial\theta} F(\underline{\beta}^{**} + \theta\underline{d})|_{\theta=0-} + \lambda\frac{\partial}{\partial\theta} f(\underline{\beta}^{**} + \theta\underline{d})|_{\theta=0-} .$$

$$\frac{\partial}{\partial\theta} S(\theta)|_{\theta=0+} = \frac{\partial}{\partial\theta} F(\underline{\beta}^{**} + \theta\underline{d}|_{\theta=0+} + \lambda\frac{\partial}{\partial\theta} f(\underline{\beta}^{**} + \theta\underline{d})|_{\theta=0+} .$$

$$\frac{\partial}{\partial\theta} F(\underline{\beta}^{**} + \theta\underline{d})|_{\theta=0-} \leq 0$$

and $$\frac{\partial}{\partial\theta} F(\underline{\beta}^{**} + \theta\underline{d})|_{\theta=0+} \geq 0$$

since $F(\underline{\beta}^{**}) \leq F(\underline{\beta})$, for all $\underline{\beta} \in R^m$. If

$$\frac{\partial}{\partial\theta} F(\underline{\beta}^{**} + \theta\underline{d})|_{\theta=0-} = 0$$

then $$F(\underline{\beta}^{**} + \theta\underline{d}) = F(\underline{\beta}^{**})$$

where $$\theta \in (\theta^o, 0) \text{ for some } \theta^o .$$

Since $f(\underline{\beta}^{**}) \leq f(\underline{\beta}^{*})$ for all $\underline{\beta}^{*} \in S^{*}$

$$f(\underline{\beta}^{**}) \leq f(\underline{\beta}^{**} + \theta\underline{d}) \quad \text{for} \quad \theta \in (\theta^{o}, 0) .$$

That is

$$\frac{\partial}{\partial\theta} f(\underline{\beta}^{**} + \theta\underline{d})\big|_{\theta=0-} \leq 0 .$$

Then, $\frac{\partial}{\partial\theta} S(\theta)\big|_{\theta=0-} \leq 0 .$

Similarly, if

$$\frac{\partial}{\partial\theta} F(\underline{\beta}^{**} + \theta\underline{d})\big|_{\theta=0+} = 0$$

then $\frac{\partial}{\partial\theta} f(\underline{\beta}^{**} + \theta\underline{d})\big|_{\theta=0+} \geq 0 .$

Hence, $\frac{\partial}{\partial\theta} S(\theta)\big|_{\theta=0+} \geq 0 .$

Let $D = \min\{-\frac{\partial}{\partial\theta} F(\underline{\beta}^{**} + \theta\underline{d})\big|_{\theta=0-} > 0 \ , \quad \frac{\partial}{\partial\theta} F(\underline{\beta}^{**} + \theta\underline{d})\big|_{\theta=0+} > 0 \ ,$ for all such $\underline{d}\}$. Note that $D > 0$ is well-defined since there is only a finite number of such $\underline{d}$. Then,

$$\frac{\partial}{\partial\theta} S(\theta)\big|_{\theta=0-} \leq - D + \lambda\frac{\partial}{\partial\theta} f(\underline{\beta}^{**} + \theta\underline{d})\big|_{\theta=0-}$$

$$\leq - D + \lambda \sum_{i=1}^{m} \operatorname{sign}(\beta_i^{**} + \theta d_i)\big|_{\theta=0-} \cdot d_i$$

$$\leq - D + \lambda \sum_{i=1}^{m} |d_i|$$

$\leq - D + \lambda \cdot M$, since we can choose $\underline{d}$ such that

$$\sum_{i=1}^{m} |d_i| \leq M \quad \text{and}$$

$\frac{\partial}{\partial\theta} S(\theta)\big|_{\theta=0-} \leq 0$, if we choose $\lambda \leq \frac{D}{M}$.

Similarly, $\frac{\partial}{\partial\theta} S(\theta)\big|_{\theta=0+} \geq D + \lambda \frac{\partial}{\partial\theta} f(\underline{\beta}^{**} + \theta\underline{d})\big|_{\theta=0+}$

$$\geq D + \lambda \sum_{i=1}^{m} \text{sign}(\beta_i^{**} + \theta d_i)\big|_{\theta=0+} \cdot d_i$$

$$\geq D - \lambda \sum_{i=1}^{m} |d_i|$$

$$\geq D - \lambda M$$

$$\geq 0 .$$

Since $\rho = \max(\frac{\partial}{\partial\theta} S(\theta)\big|_{\theta=0-} , - \frac{\partial}{\partial\theta} S(\theta)\big|_{\theta=0+}) \leq 0$ for all cases and all such $\underline{d}$ when λ is sufficiently small, the proof is complete. □

Theorem 5.3.2. If the condition of Theorem 5.3.1 holds, then $S_\lambda \leq S^{**}$ when λ is sufficiently small.

Proof: If $\underline{\beta}^{**} \in S^{**}$ satisfies the condition of Theorem 5.3.1, then $\underline{\beta}^{**} = \underline{\beta}_{\tilde{\lambda}}$ for a sufficiently small $\tilde{\lambda}$. By Lemma 5.1.3, $f(\underline{\beta}_\lambda) \geq f(\underline{\beta}_{\tilde{\lambda}})$ when $\lambda < \tilde{\lambda}$, i.e., $f(\underline{\beta}_\lambda) \leq L$ when $\lambda < \tilde{\lambda}$. By Lemma 5.1.1, $f(\underline{\beta}_\lambda) \leq L$. Thus, $f(\underline{\beta}_\lambda) = L$ when $\lambda < \tilde{\lambda}$. By Lemma 5.1.4, $F(\underline{\beta}_\lambda) \leq F(\underline{\beta}_{\tilde{\lambda}})$ when $\lambda < \tilde{\lambda}$, i.e., $\underline{\beta}_\lambda \in S^*$ when $\lambda < \tilde{\lambda}$. Thus, $S_\lambda \subseteq S^{**}$ when λ is sufficiently small. □

If there is a unique $\underline{\beta}^{**} \in S^{**}$ and the condition of Theorem 5.3.1 is satisfied, then, by Theorem 5.3.1 and Theorem 5.3.2, $\underline{\beta}_\lambda$ is identical to $\underline{\beta}^{**}$ when λ is sufficiently small. Thus, the sequence of $\underline{\beta}_\lambda$ converges to $\underline{\beta}^{**}$, the least absolute deviation estimate for the original linear model having the least ℓ_1 norm, as λ approaches zero. Further, this result can be easily extended to the augmented linear model fitting problem

$$\begin{pmatrix} X \\ \lambda I_m \end{pmatrix} \underline{\beta} \cong \begin{pmatrix} Y \\ \lambda \underline{\beta}^o \end{pmatrix}$$

where $\lambda > 0$ and $\underline{\beta}^o$ is a fixed vector in $\mathbb{R}^m$. Note that $f(\underline{\beta})$ is changed to $\sum_{j=1}^{m} |\beta_j - \beta_j^o|$. Thus, if there is a unique least absolute deviation estimate for the original linear model having the least ℓ_1 norm of the difference vector from $\underline{\beta}^o$, and residuals of the estimate vanish at m linearly independent rows in the above augmented design matrix, then the sequence of $\underline{\beta}_\lambda$ converges to the estimate as λ approaches zero.

5.4 Kuhn-Tucker Conditions in the Case $p > 1$

When $p > 1$, $\frac{\partial}{\partial \underline{\beta}} F(\underline{\beta})$ exists and is continuous for all $\underline{\beta} \in \mathbb{R}^m$. The derivative takes the form

$$\frac{\partial}{\partial \underline{\beta}} F(\underline{\beta}) = - p \sum_{i=1}^{n} |y_i - \underline{x}_i^T \underline{\beta}|^{p-1} \operatorname{sign}(y_i - \underline{x}_i^T \underline{\beta}) \underline{x}_i$$

where $\underline{x}_i^T$ is the i^{th} row in X for $1 \leq i \leq n$. Similarly,

$\frac{\partial}{\partial \underline{\beta}} f(\underline{\beta})$ exists and is continuous for all $\underline{\beta} \in \mathbb{R}^m$. This derivative can be written as

$$\frac{\partial}{\partial \underline{\beta}} f(\underline{\beta}) = p \sum_{j=1}^{m} |\beta_j|^{p-1} \operatorname{sign}(\beta_j) \underline{e}_j$$

where $\underline{e}_j$ is the j^{th} column in the identity matrix I_m for $1 \leq j \leq m$. Also, the functions $F(\underline{\beta})$ and $f(\underline{\beta})$ are convex by Theorem 2.2.2. We then can use some results in nonlinear programming to find two constrained ℓ_p estimation problems equivalent to the ℓ_p estimation problem in the augmented linear model. Further, we obtained the necessary and sufficient conditions to these equivalent problems which are known as the Kuhn-Tucker conditions. These results will be applied in the following sections.

The following problems are of interest.

Problem 1: Find a $\underline{\beta}_o \in \mathbb{R}^m$ such that

$$f(\underline{\beta}_o) - c \leq 0$$

and

$$F(\underline{\beta}_o) \leq F(\underline{\beta})$$

for all $\underline{\beta} \in \mathbb{R}^m$ such that $f(\underline{\beta}) - c \leq 0$, where c is a constant of positive number.

Problem 2: (Known as the saddle value problem.) Find a $\theta_o \geq 0$ and a $\underline{\beta}_o \in \mathbb{R}^m$ such that

$$\phi(\underline{\beta}_o, \theta) \leq \phi(\underline{\beta}_o, \theta_o) \leq \phi(\underline{\beta}, \theta_o)$$

for all $\theta \geq 0$ and $\underline{\beta} \in \mathbb{R}^m$, where $\phi(\underline{\beta}, \theta) = F(\underline{\beta}) + \theta(f(\underline{\beta}) - c)$.

Lemma 5.4.1. Problem 1 and Problem 2 are equivalent.

Proof: By Theorem 6.1 of Sposito (1975), if $\theta_o \geq 0$ and $\underline{\beta}_o$ is saddle point solution of Problem 2, then $\underline{\beta}_o$ is an optimal solution of Problem 1. Since $F(\underline{\beta})$ and $f(\underline{\beta}) - c$ are convex and there exists at least one $\underline{\beta}$ such that $f(\underline{\beta}) - c < 0$, namely $\underline{\beta} = \underline{0}$, (this is often referred to as the Slater's condition) by Theorem 6.2 of Sposito (1975), if $\underline{\beta}_o$ is an optimal solution of Problem 1, then θ_o and $\underline{\beta}_o$ is a saddle point solution of Problem 2 for some $\theta_o \geq 0$. □

Lemma 5.4.2. The Kuhn-Tucker conditions of Problem 1 and Problem 2 are:

(i) $(f(\underline{\beta}_o) - c)\theta_o = 0$,

(ii) $f(\underline{\beta}_o) - c \leq 0$,

(iii) $\theta_o \geq 0$,

(iv) $\frac{\partial}{\partial \underline{\beta}} F(\underline{\beta}_o) + \theta_o \frac{\partial}{\partial \underline{\beta}} f(\underline{\beta}_o) = 0$.

Proof: Since $\frac{\partial}{\partial \underline{\beta}} \phi(\underline{\beta}, \theta)$ and $\frac{\partial}{\partial \theta} \phi(\underline{\beta}, \theta)$ exist and are continuous at $(\underline{\beta}_o, \theta_o)$, $\phi(\underline{\beta}_o, \theta)$ is a concave function of θ, and $\phi(\underline{\beta}, \theta_o)$ is a convex function of $\underline{\beta}$, by Theorem 7.1 and Theorem 7.2 of Sposito (1975), the necessary and sufficient conditions (i.e., the Kuhn-Tucker conditions) of both problems are:

(a) $\frac{\partial}{\partial \theta} \phi(\underline{\beta}_o, \theta_o)\, \theta_o = 0$,

(b) $\frac{\partial}{\partial \theta} \phi(\underline{\beta}_o, \theta_o) \leq 0$,

(c) $\theta_o \geq 0$,

(d) $\frac{\partial}{\partial \underline{\beta}} \phi(\underline{\beta}_o, \theta_o) = 0$.

In other words, the Kuhn-Tucker conditions of Problem 1 and Problem 2 are:

(i) $(f(\underline{\beta}_o) - c)\, \theta_o = 0$,

(ii) $f(\underline{\beta}_o) - c \leq 0$,

(iii) $\theta_o \geq 0$,

(iv) $\frac{\partial}{\partial \underline{\beta}} F(\underline{\beta}_o) + \theta_o \frac{\partial}{\partial \underline{\beta}} f(\underline{\beta}_o) = 0$,

note that, if $\theta_o > 0$, then $f(\underline{\beta}_o) - c = 0$ by the condition (i). □

Since the design matrix in the augmented linear model is full-rank, as indicated in Section 2.3, there is a unique optimal $\underline{\beta}_\lambda \in \mathbb{R}^m$. The necessary and sufficient condition of $\underline{\beta}_\lambda$ can be obtained by setting the gradient vector zero vector, i.e.,

$$\frac{\partial}{\partial \underline{\beta}} F(\underline{\beta}_\lambda) + \lambda \frac{\partial}{\partial \underline{\beta}} F(\underline{\beta}_\lambda) = \underline{0} .$$

We then choose $\underline{\beta}_o = \underline{\beta}_\lambda$ and $\theta_o = \lambda$ such that the Kuhn-Tucker conditions are satisfied and $c = f(\underline{\beta}_\lambda)$. Note that,

$$\phi(\underline{\beta}_\lambda, \theta_o) = F(\underline{\beta}_\lambda) + \theta_o(f(\underline{\beta}_\lambda) - c) = F(\underline{\beta}_\lambda)$$

and $\phi(\underline{\beta}_\lambda, \theta) = F(\underline{\beta}_\lambda) = \phi(\underline{\beta}_\lambda, \theta_o)$, for all $\theta \geq 0$,

in Problem 2. Hence, Problem 2 can be stated as, given $\lambda > 0$, to find a $\underline{\beta}_\lambda \in \mathbb{R}^m$ such that

$$F(\underline{\beta}_\lambda) + \lambda f(\underline{\beta}_\lambda) \leq F(\underline{\beta}) + \lambda f(\underline{\beta}) \text{ , for all } \underline{\beta} \in \mathbb{R}^m .$$

which is the ℓ_p estimation problem in the augmented linear model. The equivalent Problem 1 is to find a $\underline{\beta}_\lambda$ such that $F(\underline{\beta}_\lambda) \leq F(\underline{\beta})$, for all $\underline{\beta} \in \mathbb{R}^m$ such that $f(\underline{\beta}) \leq f(\underline{\beta}_\lambda)$, which is a constrained ℓ_p estimation problem.

Similarly, if we consider the following two problems:

Problem 1': Find a $\underline{\beta}_o \in \mathbb{R}^m$ such that

$$F(\underline{\beta}_o) - \rho \leq 0$$

and

$$f(\underline{\beta}_o) \leq f(\underline{\beta})$$

for all $\underline{\beta} \in \mathbb{R}^m$ such that $F(\underline{\beta}) - \rho \leq 0$, where ρ is a constant of a number greater than $F(\underline{\beta}^*)$.

Problem 2': Find a $w_o \geq 0$ and a $\underline{\beta}_o \in \mathbb{R}^m$ such that

$$\psi(\underline{\beta}_o, w) \leq \psi(\underline{\beta}_o, w_o) \leq \psi(\underline{\beta}, w_o)$$

for all $w \geq 0$ and $\underline{\beta} \in \mathbb{R}^m$, where $\psi(\underline{\beta}, w) = f(\underline{\beta}) + w(F(\underline{\beta}) - \rho)$. Then by Theorem 6.1 and Theorem 6.2 of Sposito (1975), we have the following lemma.

Lemma 5.4.3. Problem 1' and Problem 2' are equivalent.

Also, by Theorem 7.1 and Theorem 7.2 of Sposito (1975), we have the following lemma.

Lemma 5.4.4. The Kuhn-Tucker conditions of Problem 1' and Problem 2' are:

(i') $(F(\underline{\beta}_o) - \rho)w_o = 0$,

(ii') $F(\underline{\beta}_o) - \rho \leq 0$,

(iii') $w_o \geq 0$,

(iv') $\frac{\partial}{\partial \underline{\beta}} f(\underline{\beta}_o) + w_o \frac{\partial}{\partial \underline{\beta}} F(\underline{\beta}_o) = 0$.

We then choose $\underline{\beta}_o = \underline{\beta}_\lambda$ and $w_o = \frac{1}{\lambda}$ such that the Kuhn-Tucker conditions are satisfied and $\rho = F(\underline{\beta}_\lambda)$. Then, Problem 2' is the ℓ_p estimation problem in the augmented linear model. The equivalent Problem 1' is to find a $\underline{\beta}_\lambda \in \mathbb{R}^m$ such that $f(\underline{\beta}_\lambda) \leq f(\underline{\beta})$ for all $\underline{\beta} \in \mathbb{R}^m$ such that $F(\underline{\beta}) \leq F(\underline{\beta}_\lambda)$.

Therefore, $F(\underline{\beta}_\lambda) + f(\underline{\beta}_\lambda) \leq F(\underline{\beta}) + \lambda f(\underline{\beta})$ for all $\underline{\beta} \in \mathbb{R}^m$, if and only if, $F(\underline{\beta}_\lambda) \leq F(\underline{\beta})$ for all $\underline{\beta} \in \mathbb{R}^m$ such that $f(\underline{\beta}) \leq f(\underline{\beta}_\lambda)$, and, if and only if, $f(\underline{\beta}_\lambda) \leq f(\underline{\beta})$ for all $\underline{\beta} \in \mathbb{R}^m$ such that $F(\underline{\beta}) \leq F(\underline{\beta}_\lambda)$. Moreover, $\underline{\beta}_\lambda$ is unique as indicated earlier, hence, we have the following theorem.

Theorem 5.4.1. $F(\underline{\beta}_\lambda) < F(\underline{\beta})$ for all $\underline{\beta} \in \mathbb{R}^m$ such that $\underline{\beta} \neq \underline{\beta}_\lambda$ and $f(\underline{\beta}) \leq f(\underline{\beta}_\lambda)$. Also, $f(\underline{\beta}_\lambda) < f(\underline{\beta})$ for all $\underline{\beta} \in \mathbb{R}^m$ such that $\underline{\beta} \neq \underline{\beta}_\lambda$ and $F(\underline{\beta}) \leq F(\underline{\beta}_\lambda)$.

5.5 Properties of $\underline{\beta}_\lambda$ in the Case $p > 1$

Since $\underline{\beta}_\lambda$ is unique for every $\lambda > 0$, $F(\underline{\beta}_\lambda)$ and $f(\underline{\beta}_\lambda)$ are functions of λ , denoted as $\bar{F}(\lambda)$ and $\bar{f}(\lambda)$, respectively. We will characterize functions $\bar{F}(\lambda)$ and $\bar{f}(\lambda)$ and will obtain the limiting $\underline{\beta}_\lambda$ as λ approach zero. We now proceed to prove some lemmas which are needed in the following theorem.

Since $\frac{\partial}{\partial \underline{\beta}} F(\underline{\beta}^*) = \underline{0}$ and $\frac{\partial}{\partial \underline{\beta}} F(\underline{\beta}_\lambda) + \lambda \frac{\partial}{\partial \underline{\beta}} f(\underline{\beta}_\lambda) = \underline{0}$, if $\underline{\beta}^* = \underline{\beta}_\lambda$ then $\frac{\partial}{\partial \underline{\beta}} f(\underline{\beta}^*) = \underline{0}$, which implies $\underline{\beta}^* = \underline{0}$ and this is not true in general. Thus, $\underline{\beta}^* \neq \underline{\beta}_\lambda$ in most cases. Then, by Theorem 5.4.1 and Lemma 5.1.2, we have the following lemma.

Lemma 5.5.1. $f(\underline{\beta}_\lambda) < f(\underline{\beta}^*)$ and $f(\underline{\beta}_\lambda) < L$.

By Lemma 5.1.2, we have the following lemma.

Lemma 5.5.2. $F(\underline{\beta}_\lambda) > M$.

Since $\frac{\partial}{\partial \underline{\beta}} F(\underline{\beta}_{\lambda_1}) + \lambda_1 \frac{\partial}{\partial \underline{\beta}} f(\underline{\beta}_{\lambda_1}) = \underline{0}$ and $\frac{\partial}{\partial \underline{\beta}} F(\underline{\beta}_{\lambda_2}) + \lambda_2 \frac{\partial}{\partial \underline{\beta}} f(\underline{\beta}_{\lambda_2}) = 0$, if $\underline{\beta}_{\lambda_1} = \underline{\beta}_{\lambda_2}$ then $\lambda_1 = \lambda_2$. Hence, $\underline{\beta}_{\lambda_1} \neq \underline{\beta}_{\lambda_2}$ when $\lambda_1 \neq \lambda_2$. With the strict inequalities in Lemma 5.1.3 and Lemma 5.1.4, we have the following lemma.

Lemma 5.5.3. $f(\underline{\beta}_{\lambda_1}) < f(\underline{\beta}_{\lambda_2})$ when $\lambda_1 > \lambda_2$.

Proof: Since $\underline{\beta}_{\lambda_1}$ is unique, we have

$$F(\underline{\beta}_{\lambda_1}) + \lambda_1 f(\underline{\beta}_{\lambda_1}) < F(\underline{\beta}_{\lambda_2}) + \lambda_1 f(\underline{\beta}_{\lambda_2})$$

since $\underline{\beta}_{\lambda_1} \neq \underline{\beta}_{\lambda_2}$. Similarly, we have

$$F(\underline{\beta}_{\lambda_2}) + \lambda_2 f(\underline{\beta}_{\lambda_2}) < F(\underline{\beta}_{\lambda_1}) + \lambda_2 f(\underline{\beta}_{\lambda_1}) .$$

Hence,

$$(\lambda_1 - \lambda_2) F(\underline{\beta}_{\lambda_1}) < (\lambda_1 - \lambda_2) f(\underline{\beta}_{\lambda_2}) .$$

Since $\lambda_1 - \lambda_2 > 0$, we have $f(\underline{\beta}_{\lambda_1}) < f(\underline{\beta}_{\lambda_2})$. □

Lemma 5.5.4. $F(\underline{\beta}_{\lambda_1}) > F(\underline{\beta}_{\lambda_2})$ when $\lambda_1 > \lambda_2$.

Proof: Since $\underline{\beta}_{\lambda_1}$ is unique, we have

$$\frac{1}{\lambda_1} F(\underline{\beta}_{\lambda_1}) + f(\underline{\beta}_{\lambda_1}) < \frac{1}{\lambda_1} F(\underline{\beta}_{\lambda_2}) + f(\underline{\beta}_{\lambda_2}) .$$

Similarly,

$$\frac{1}{\lambda_2} F(\underline{\beta}_{\lambda_2}) + f(\underline{\beta}_{\lambda_2}) < \frac{1}{\lambda_2} F(\underline{\beta}_{\lambda_1}) + f(\underline{\beta}_{\lambda_1}) .$$

Then,

$$\left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1}\right) F(\underline{\beta}_{\lambda_2}) < \left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1}\right) F(\underline{\beta}_{\lambda_1}) .$$

Since $\frac{1}{\lambda_2} - \frac{1}{\lambda_1} > 0$, we have $F(\underline{\beta}_{\lambda_1}) > F(\underline{\beta}_{\lambda_2})$. □

Now, suppose that $\underline{\beta}_{\lambda} = \underline{0}$. Then,

$$\frac{\partial}{\partial \underline{\beta}} F(\underline{0}) + \lambda \frac{\partial}{\partial \underline{\beta}} f(\underline{0}) = 0$$

implies $\frac{\partial}{\partial \underline{\beta}} F(\underline{0}) = \underline{0}$, since $\frac{\partial}{\partial \underline{\beta}} f(\underline{0}) = \underline{0}$. Hence, $\underline{\beta}^* = \underline{0}$, which is not true in general. Therefore, $\underline{\beta}_{\lambda} \neq \underline{0}$. Then, we have the following lemma which follows from the definition of f .

Lemma 5.5.5. $f(\underline{\beta}_{\lambda}) > 0$.

By Theorem 5.4.1 and Lemma 5.5.5, we have the following lemma.

Lemma 5.5.6. $F(\underline{\beta}_{\lambda}) < F(\underline{0})$.

Now we are ready to prove an important result as given in the following theorem.

Theorem 5.5.1. $\bar{f}(\lambda)$ is a strictly decreasing, continuous, and onto function of λ from $(0, \infty)$ to $(0, L)$, such that $\lim_{\lambda \to 0} \bar{f}(\lambda) = L$ and $\lim_{\lambda \to \infty} \bar{f}(\lambda) = 0$.

Proof: By Lemma 5.5.1 and Lemma 5.5.5, we have $0 < \bar{f}(\lambda) < L$ for all $\lambda > 0$. Thus, $\bar{f}(\lambda)$ is a function from $(0,\infty)$ to $(0, L)$. By Lemma 5.5.3, $\bar{f}(\lambda)$ is strictly decreasing.

We will now prove that $\bar{f}(\lambda)$ is an onto function. Let $0 < c < L$, and let $B = \{\underline{\beta} | f(\underline{\beta}) \leq c\}$. Since $F(\underline{\beta})$ is continuous and B is closed and bounded, by Lemma 2.3.1, there exists a $\underline{\beta}_o$ B such that $F(\underline{\beta}_o) \leq F(\underline{\beta})$ for all $\underline{\beta} \in B$. By Lemma 5.4.1, there is an equivalent saddle point problem, i.e., there exists a $\theta_o \geq 0$ such that

$$\phi(\underline{\beta}_o, \theta) \leq \phi(\underline{\beta}_o, \theta_o) \leq \phi(\underline{\beta}, \theta_o)$$

for all $\theta \geq 0$ and $\underline{\beta} \in \mathbb{R}^m$, where $\phi(\underline{\beta}, \theta) = F(\underline{\beta}) + \theta(f(\underline{\beta}) - c)$. Thus, there exists a $\theta_o \geq 0$ such that

$$F(\underline{\beta}_o) + \theta_o f(\underline{\beta}_o) \leq F(\underline{\beta}) + \theta_o f(\underline{\beta})$$

for all $\underline{\beta} \in \mathbb{R}^m$. Hence, $\underline{\beta}_o = \underline{\beta}_\lambda$, where $\lambda = \theta_o$, by the definition of $\underline{\beta}_\lambda$. $\theta_o > 0$, since if $\theta_o = 0$, then $\underline{\beta}_o = \underline{\beta}^*$, but $f(\underline{\beta}_o) \leq c < L \leq f(\underline{\beta}^*)$, which leads to a contradiction. Then, by the condition (i) of Lemma 5.4.2, $c = f(\underline{\beta}_\lambda)$, where $\lambda = \theta_o$. Thus, $c = \bar{f}(\theta_o)$ for some $\theta_o > 0$. Hence, $\bar{f}(\lambda)$ is an onto function.

We will now prove that $\bar{f}(\lambda)$ is left-continuous on $(0, \infty]$. We define $\bar{f}(\infty) \equiv 0$. Let $\lambda_o \in (0, \infty]$. Given $\varepsilon > 0$ sufficiently small, by the fact that $\bar{f}(\lambda)$ is an onto function, there exists a $\lambda_\varepsilon > 0$ such that $\bar{f}(\lambda_\varepsilon) = \bar{f}(\lambda_o) + \varepsilon$. Then, by Lemma 5.5.3,

$$f(\lambda_o) + \varepsilon > \bar{f}(\lambda) > \bar{f}(\lambda_o)$$

when $\lambda_\varepsilon < \lambda < \lambda_o$. Thus, $\bar{f}(\lambda)$ is left-continuous at λ_o. As $\lambda_o \to \infty$, we have $\lim_{\lambda\to\infty} \bar{f}(\lambda) = 0$. Further, $\bar{f}(\lambda)$ is right-continuous on $[0, \infty)$. We define $\bar{f}(0) \equiv L$. Let $\lambda_o \in [0, \infty)$. Given $\varepsilon' > 0$ sufficiently small, by the fact that $\bar{f}(\lambda)$ is an onto function, there exists a $\lambda_{\varepsilon'} > 0$ such that $\bar{f}(\lambda_{\varepsilon'}) = \bar{f}(\lambda_o) - \varepsilon'$. Then, by Lemma 5.5.3,

$$\bar{f}(\lambda_o) > \bar{f}(\lambda) > \bar{f}(\lambda_o) - \varepsilon'$$

when $\lambda_o < \lambda < \lambda_{\varepsilon'}$. Thus, $\bar{f}(\lambda)$ is right-continuous at λ_o. For $\lambda_o = 0$, we have $\lim_{\lambda\to 0} \bar{f}(\lambda) = \bar{f}(0) = L$. Since $\bar{f}(\lambda)$ is left-continuous and right-continuous at $\lambda_o \in (0, \infty)$, $\bar{f}(\lambda)$ is continuous on $(0, \infty)$. □

Theorem 5.5.2. $\bar{F}(\lambda)$ is a strictly increasing, continuous, and onto function of λ from $(0, \infty)$ to $(M, F(\underline{0}))$, such that $\lim_{\lambda\to 0} \bar{F}(\lambda) = M$ and $\lim_{\lambda\to\infty} \bar{F}(\lambda) = F(\underline{0})$.

Proof: By Lemma 5.5.2 and Lemma 5.5.6, we have $M < \bar{F}(\lambda) < F(\underline{0})$ for all $\lambda > 0$. Thus, $\bar{F}(\lambda)$ is a function from $(0, \infty)$ to $(M, F(\underline{0}))$. By Lemma 5.5.4, $\bar{F}(\lambda)$ is strictly increasing.

We will now prove that $\bar{F}(\lambda)$ is an onto function. Let $M < \rho < F(\underline{0})$, and $B = \{\underline{\beta} \in \mathbb{R}^m | F(\underline{\beta}) \leq \rho\}$. $F(\underline{\beta}^*) = M \leq \rho$ implies that B is not empty. Also, B is closed. The design matrix is $\lambda^{\frac{1}{p}} I_m$ in this case, hence, it is full-rank. By Lemma 2.3.3, there exists a $\underline{\beta}_o \in B$ such that $f(\underline{\beta}_o) \leq f(\underline{\beta})$ for all $\underline{\beta} \in B$. By Lemma 5.4.3, there is an equivalent saddle point problem, i.e., there exists a $w_o \geq 0$ such that

$$\psi(\underline{\beta}_o, w) \leq \psi(\underline{\beta}_o, w_o) \leq \psi(\underline{\beta}, w_o)$$

for all $w \geq 0$ and $\underline{\beta} \in \mathbb{R}^m$, where $\psi(\underline{\beta}, w) = f(\underline{\beta}) + w(F(\underline{\beta}) - \rho)$. Thus, there exists a $w_o \geq 0$ such that

$$f(\underline{\beta}_o) + w_o F(\underline{\beta}_o) \leq f(\underline{\beta}) + w_o F(\underline{\beta})$$

for all $\underline{\beta} \in \mathbb{R}^m$. Hence, $\underline{\beta}_o = \underline{\beta}_\lambda$, where $\lambda = \frac{1}{w_o}$, by the definition of $\underline{\beta}_\lambda$. $w_o > 0$, since if $w_o = 0$, then $\underline{\beta}_o = \underline{0}$, but $F(\underline{\beta}_o) \leq \rho < F(\underline{0})$, which leads to a contradiction. Then, by the condition (i') of Lemma 5.4.4, $\rho = F(\underline{\beta}_\lambda)$, where $\lambda = \frac{1}{w_o}$. Thus, $\rho = \bar{F}(\frac{1}{w_o})$ for some $w_o > 0$. Hence, $\bar{F}(\lambda)$ is an onto function.

We will now prove that $\bar{F}(\lambda)$ is left-continuous on $(0, \infty]$. We define $\bar{F}(\infty) = F(\underline{0})$. Let $\lambda_o \in (0, \infty]$. Given $\varepsilon > 0$ sufficiently small, by the fact that $\bar{F}(\lambda)$ is an onto function, there exists a $\lambda_\varepsilon > 0$ such that $\bar{F}(\lambda_\varepsilon) = \bar{F}(\lambda_o) - \varepsilon$. Then, by Lemma 5.5.4,

$$\bar{F}(\lambda_o) - \varepsilon < \bar{F}(\lambda) < \bar{F}(\lambda_o), \quad \text{when } \lambda_\varepsilon < \lambda < \lambda_o.$$

Thus, $\bar{F}(\lambda)$ is left-continuous at λ_o. As $\lambda_o \to \infty$. We have

$\lim_{\lambda\to\infty} \bar{F}(\lambda) = \bar{F}(\infty) = F(0)$. Further, $\bar{F}(\lambda)$ is right-continuous on $[0, \infty)$. We define $\bar{F}(0) \equiv M$ and let $\lambda_o \in [0, \infty)$. Given $\varepsilon' > 0$ sufficiently small, by the fact that $\bar{F}(\lambda)$ is an onto function, there exists a $\lambda_{\varepsilon'} > 0$ such that $\bar{F}(\lambda_{\varepsilon'}) = \bar{F}(\lambda_o) + \varepsilon'$. Then, by Lemma 5.5.4,

$$\bar{F}(\lambda_o) < \bar{F}(\lambda) < \bar{F}(\lambda_o) + \varepsilon' \text{ , when } \lambda_o < \lambda < \lambda_{\varepsilon'} .$$

Hence, $\bar{F}(\lambda)$ is right-continuous at λ_o . For $\lambda_o = 0$, we have $\lim_{\lambda\to 0} \bar{F}(\lambda) = \bar{F}(0) = M$. Since $\bar{F}(\lambda)$ is left-continuous and right-continuous at $\lambda_o \in (0, \infty)$, $\bar{F}(\lambda)$ is continuous on $(0, \infty)$. □

As indicated in Section 2.3, the ℓ_p estimate is not unique when X is not full-rank. But there exists an ℓ_p estimate having the least ℓ_p norm, as in the case of $p = 2$, to which the sequence of $\underline{\beta}_\lambda$ converges as λ approaches zero. We assume X is not full-rank in the rest of the section and derive proofs of these results.

Now, we consider any sequence $Q = \{\underline{\beta}_{\lambda_i}\}_{i=1}^{\infty}$ such that $\lambda_i \to 0$ when $i \to \infty$. As indicated earlier $\underline{\beta}_{\lambda_i} \neq \underline{\beta}_{\lambda_j}$ if $\lambda_i \neq \lambda_j$. Also, by Lemma 5.5.1, $f(\underline{\beta}_{\lambda_i}) < L$, i.e., $\|\beta_{\lambda_i}\|_p < L^p$, for all i . Therefore, Q is a bounded infinite sequence in $\mathbb{R}^m$.

Lemma 5.5.7. Let Q' be a subsequence of Q , then there exists a $\underline{\beta}^{**} \in \mathbb{R}^m$ and a further subsequence $Q'' = \{\underline{\beta}_{\lambda_i''}\}_{i=1}^{\infty}$ of the subsequence Q' where $\lambda_i'' \to 0$ when $i \to \infty$, such that $\lim_{i\to\infty} \underline{\beta}_{\lambda_i''} = \underline{\beta}^{**}$, where $F(\underline{\beta}^{**}) = M$ and $f(\underline{\beta}^{**}) = L$.

Proof: Since Q' is a bounded infinite subset in $\mathbb{R}^m$, by the Bolzano-Weierstrass theorem, there exists a cluster vector $\underline{\beta}^{**} \in \mathbb{R}^m$. Then, there exists a further subsequence $Q'' = \{\underline{\beta}_{\lambda_i''}\}_{i=1}^{\infty}$ of the subsequence Q', where $\lambda_i'' \to 0$ when $i \to \infty$, such that $\lim_{i\to\infty} \underline{\beta}_{\lambda_i''} = \underline{\beta}^{**}$. Since $f(\underline{\beta})$ and $F(\underline{\beta})$ are continuous, we have $\lim_{i\to\infty} f(\underline{\beta}_{\lambda_i''}) = f(\underline{\beta}^{**})$ and $\lim_{i\to\infty} F(\underline{\beta}_{\lambda_i''}) = F(\underline{\beta}^{**})$. By Theorem 5.5.1 and Theorem 5.5.2, we have $\lim_{i\to\infty} f(\underline{\beta}_{\lambda_i''}) = L$ and $\lim_{i\to\infty} F(\underline{\beta}_{\lambda_i''}) = M$. Thus, $f(\underline{\beta}^{**}) = L$ and $F(\underline{\beta}^{**}) = M$. □

The following lemma shows that the two conditions which $\underline{\beta}^{**}$ satisfies can uniquely determine $\underline{\beta}^{**} \in \mathbb{R}^m$.

Lemma 5.5.8. There is only one $\underline{\beta}^{**} \in \mathbb{R}^m$ such that $F(\underline{\beta}^{**}) = M$ and $f(\underline{\beta}^{**}) = L$.

Proof: We suppose that there exists another $\underline{\beta}^{**'} \in R^m$ such that $F(\underline{\beta}^{**'}) = M$ and $f(\underline{\beta}^{**'}) = L$. Then,

$$f(\alpha\underline{\beta}^{**} + (1-\alpha)\underline{\beta}^{**'}) < \alpha f(\underline{\beta}^{**}) + (1-\alpha)f(\underline{\beta}^{**'}) .$$

Hence, $0 < \alpha < 1$, since $F(\underline{\beta})$ is strictly convex by Theorem 2.2.2.

$$f(\alpha\underline{\beta}^{**} + (1-\alpha)\underline{\beta}^{**'}) < L .$$

However,

$$F(\alpha\underline{\beta}^{**} + (1-\alpha)\underline{\beta}^{**'}) \leq \alpha F(\underline{\beta}^{**}) + (1-\alpha)F(\underline{\beta}^{**'})$$

since $F(\underline{\beta})$ is convex. Hence,

$$F(\alpha\underline{\beta}^{**} + (1-\alpha)\underline{\beta}^{**'}) \leq M ,$$

i.e., $\alpha\underline{\beta}^{**} + (1-\alpha)\underline{\beta}^{**'} \in S^{*}$. Therefore,

$$f(\alpha\underline{\beta}^{**} + (1-\alpha)\underline{\beta}^{**'}) \geq L$$

which leads to a contradiction. Hence, there is only one $\underline{\beta}^{**} \in \mathbb{R}^{m}$ such that $F(\underline{\beta}^{**}) = M$ and $f(\underline{\beta}^{**}) = L$. □

The following theorem shows that, for $p > 1$, the sequence of $\underline{\beta}_{\lambda}$ converges to $\underline{\beta}^{**}$, the ℓ_p estimate in the original model having the least ℓ_p norm, as λ approaches zero.

Theorem 5.5.3. Given any sequence $Q = \{\underline{\beta}_{\lambda_i}\}_{i=1}^{\infty}$ such that $\lambda_i \to 0$ when $i \to \infty$, the sequence converges to $\underline{\beta}^{**}$.

Proof: By Lemma 5.5.7, for any subsequence Q' , there is a further subsequence Q'' of the subsequence Q' which converges to $\underline{\beta}^{**}$. By Lemma 5.5.8, there is only one such $\underline{\beta}^{**}$. Hence, the sequence Q converges to $\underline{\beta}^{**}$. □

5.6 Discussion on Generalization and Application in the Case $p > 1$

We can use the same technique in Section 5.5 to identify vectors other than $\underline{\beta}^{**}$ in S^{*} when X is not full-rank. Let

$$F(\underline{\beta}) = \sum_{i=1}^{n} |y_i - \underline{x}_i^{T}\underline{\beta}|^{p}$$

as in Section 5.5, but let

$$f(\underline{\beta}) = \sum_{j=1}^{m} |\beta_j - \beta_j^{o}|^{p}$$

where $\underline{\beta}^o$ is a given vector in $\mathbb{R}^m$. Let $\underline{\alpha} = \underline{\beta} - \underline{\beta}^o$. Then,

$$F(\underline{\beta}) = \sum_{i=1}^{n} |y_i - \underline{x}_i^T(\underline{\alpha} + \underline{\beta}^o)|^p$$

$$= \sum_{i=1}^{n} |z_i - \underline{x}_i^T\underline{\alpha}|^p .$$

denoted as $G(\underline{\alpha})$, where $z_i = y_i - \underline{x}_i^T\underline{\beta}^o$. Also,

$$f(\underline{\beta}) = \sum_{j=1}^{m} |\alpha_j|^p$$

denoted as $g(\underline{\alpha})$. Let $\underline{\alpha}_\lambda \in \mathbb{R}^m$ such that

$$G(\underline{\alpha}_\lambda) + \lambda g(\underline{\alpha}_\lambda) \leq G(\underline{\alpha}) + \lambda g(\underline{\alpha}) \text{ , for all } \underline{\alpha} \in \mathbb{R}^m$$

where $\lambda > 0$. Note that, $\underline{\alpha}_\lambda + \underline{\beta}^o = \underline{\beta}_\lambda$, where $\underline{\beta}_\lambda$ is defined as in Section 5.1 with the $f(\underline{\beta})$ in this section.

By Theorem 5.5.3, $\lim_{\lambda \to 0} \underline{\alpha}_\lambda = \underline{\alpha}^{**}$, where $\underline{\alpha}^{**} \in \mathbb{R}^m$ is unique, such that $G(\underline{\alpha}^{**}) \leq G(\underline{\alpha})$ for all $\underline{\alpha} \in \mathbb{R}^m$, and $g(\underline{\alpha}^{**}) \leq g(\underline{\alpha}^*)$ for all $\underline{\alpha}^* \in \mathbb{R}^m$, such that $G(\underline{\alpha}^*) \leq G(\underline{\alpha})$ for all $\underline{\alpha} \in \mathbb{R}^m$. Therefore, $\lim_{\lambda \to 0} \underline{\beta}_\lambda = \underline{\alpha}^{**} + \underline{\beta}^o$, $F(\underline{\alpha}^{**} + \underline{\beta}^o) \leq F(\underline{\beta})$ for all $\underline{\beta} \in \mathbb{R}^m$, and $f(\underline{\alpha}^{**} + \underline{\beta}^o) \leq f(\underline{\beta}^*)$ for all $\underline{\beta}^* \in S^*$, where S^* is defined as in Section 5.1. Hence, the sequence of $\underline{\beta}_\lambda$ converges to the ℓ_p estimate for the original model having the least ℓ_p norm of the difference vector from the given $\underline{\beta}^o$.

In case of $p = 2$, $\lim_{\lambda \to 0} \underline{\alpha}_\lambda = X^+\underline{z} = \underline{\alpha}^{**}$, where X^+ is the

Moore-Penrose inverse of X and $\underline{z} = \underline{y} - X\underline{\beta}^{o}$. Therefore, the least squares estimate for the original model having the least Euclidean distance from the given $\underline{\beta}^{o}$ is $X^{+}\underline{y} - X^{+}X\underline{\beta}^{o} + \underline{\beta}^{o}$. Let

$$X = Q\begin{pmatrix} R_1 & 0 \\ 0 & 0 \end{pmatrix} U^{T}$$

and

$$X^{+} = U\begin{pmatrix} R_1^{-1} & 0 \\ 0 & 0 \end{pmatrix} Q^{T},$$

as in Section 5.2. Then,

$$\begin{aligned} X^{+}X &= U\begin{pmatrix} R_1^{-1} & 0 \\ 0 & 0 \end{pmatrix} Q^{T}Q \begin{pmatrix} R_1 & 0 \\ 0 & 0 \end{pmatrix} U^{T} \\ &= U\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} U^{T} \\ &= (U_1U_2)\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} U_1^{T} \\ U_2^{T} \end{pmatrix} \\ &= U_1U_1^{T} \end{aligned}$$

where $U = (U_1U_2)$. Also, $X^{+}\underline{y} = U_1\underline{a}_1^{*}$, where $\underline{a}_1^{*} = R_1^{-1}Q_1^{T}\underline{y}$ and $Q = (Q_1Q_2)$, as indicated in Section 5.2. Hence, $X^{+}\underline{y} - X^{+}X\underline{\beta}^{o} + \underline{\beta}^{o} = U_1\underline{a}_1^{*} - U_1U_1^{T}\underline{\beta}^{o} + \underline{\beta}^{o}$. As indicated in Section 2.1, $\underline{\beta}^{o}$ is a least squares estimate for the original linear model if and only if, $U_1^{T}\underline{\beta}^{o} = \underline{a}_1^{*}$. Thus, $X^{+}\underline{y} - X^{+}X\underline{\beta}^{o} + \underline{\beta}^{o} = \underline{\beta}^{o}$ when $\underline{\beta}^{o}$ is a least squares estimate for the original linear model.

Further, let $F(\underline{\beta})$ be a convex function of $\underline{\beta}$ in $\mathbb{R}^m$ such that $\frac{\partial}{\partial \underline{\beta}} F(\underline{\beta})$ exists and is continuous, and let $f(\underline{\beta})$ be a strictly convex function of $\underline{\beta}$ in $\mathbb{R}^m$ such that $\frac{\partial}{\partial \underline{\beta}} f(\underline{\beta})$ exists and is continuous. Let $D_c = \{\underline{\beta} \in \mathbb{R}^m | f(\underline{\beta}) \leq c\}$, where $c \in \mathbb{R}$. We assume that D_c is bounded for all $c \in \mathbb{R}$. Note that D_c is closed, since f is continuous implies that the complement subset of D_c, $\{\underline{\beta} \in \mathbb{R}^m | f(\underline{\beta}) > c\}$, is open. Hence, by Lemma 2.3.1, if $D_c \neq \emptyset$, there exists a $\underline{\beta}^o \in D_c$ such that $f(\underline{\beta}^o) \leq f(\underline{\beta})$ for all $\underline{\beta} \in D_c$. Therefore, $f(\underline{\beta}^o) \leq f(\underline{\beta})$ for all $\underline{\beta} \in \mathbb{R}^m$. Moreover, $\underline{\beta}^o$ is unique since f is strictly convex, and $\frac{\partial}{\partial \underline{\beta}} f(\underline{\beta}^o) = \underline{0}$. Let $S^* = \{\underline{\beta}^* \in \mathbb{R}^m | F(\underline{\beta}^*) \leq F(\underline{\beta})$ for all $\underline{\beta} \in \mathbb{R}^m\}$. We assume that $S^* = \emptyset$ and $\underline{\beta}^o \notin S^*$. Note that $\frac{\partial}{\partial \underline{\beta}} F(\underline{\beta}^*) = \underline{0}$. Also, we assume that there is a $\underline{\beta}_\lambda \in \mathbb{R}^m$ such that $F(\underline{\beta}_\lambda) + \lambda f(\underline{\beta}_\lambda) \leq F(\underline{\beta}) + \lambda f(\underline{\beta})$ for all $\underline{\beta} \in \mathbb{R}^m$, where $\lambda > 0$. Note that $\underline{\beta}_\lambda$ is unique since $F + \lambda f$ is strictly convex, and $\frac{\partial}{\partial \underline{\beta}} F(\underline{\beta}_\lambda) + \lambda \frac{\partial}{\partial \underline{\beta}} f(\underline{\beta}_\lambda) = \underline{0}$. Let $B_\rho = \{\underline{\beta} \in \mathbb{R}^m | F(\underline{\beta}) \leq \rho\}$. Let $M = F(\underline{\beta}^*)$, where $\underline{\beta}^* \in S^*$, and let $L = \inf f\{\underline{\beta}^* | \underline{\beta}^* \in S^*\}$. We assume that, for any $\rho > M$, there exists a $\tilde{\underline{\beta}} \in B_\rho$ such that $f(\tilde{\underline{\beta}}) \leq f(\underline{\beta})$ for all $\underline{\beta} \in B_\rho$. Now we have all conditions needed for f and F in Section 5.5. Let $\bar{f}(\lambda) = f(\underline{\beta}_\lambda)$ and $\bar{F}(\lambda) = F(\underline{\beta}_\lambda)$, then we have the following theorem.

Theorem 5.6.1. $\bar{f}(\lambda)$ is a strictly decreasing, continuous, and onto function of λ from $(0, \infty)$ to $(f(\underline{\beta}^o), L)$, such that $\lim_{\lambda \to 0} \bar{f}(\lambda) = L$ and $\lim_{\lambda \to \infty} \bar{f}(\lambda) = f(\underline{\beta}^o)$. $\bar{F}(\lambda)$ is a strictly increasing, continuous, and onto function of λ from $(0, \infty)$ to $(M, F(\underline{\beta}^o))$, such that $\lim_{\lambda \to 0} \bar{F}(\lambda) = M$

and $\lim_{\lambda\to\infty} \bar{F}(\lambda) = F(\underline{\beta}^o)$. Further, when S^* contains more than one vector in $\mathbb{R}^m$, there exists a unique $\underline{\beta}^{**} \in S^*$ such that $f(\underline{\beta}^{**}) \leq f(\underline{\beta}^*)$ for all $\underline{\beta}^* \in S^*$, and $\lim_{\lambda\to 0} \underline{\beta}_\lambda = \underline{\beta}^{**}$.

We now apply Theorem 5.6.1 in ℓ_p estimation under linear equality restrictions. Let us consider the problem of finding $\underline{\beta}^{**} \in \mathbb{R}^m$ such that $A\underline{\beta}^{**} = \underline{b}$ and $\sum_{i=1}^{n} |y_i - \underline{x}_i^T\underline{\beta}^{**}|^p \leq \sum_{i=1}^{n} |y_i - \underline{x}_i^T\underline{\beta}^*|^p$, $p > 1$, for all $\underline{\beta}^* \in \mathbb{R}^m$ such that $A\underline{\beta}^* = \underline{b}$, where $X = \begin{pmatrix} \underline{x}_1^T \\ \underline{x}_2^T \\ \vdots \\ \underline{x}_n^T \end{pmatrix}$ and $A = \begin{pmatrix} \underline{a}_1^T \\ \underline{a}_2^T \\ \vdots \\ \underline{a}_t^T \end{pmatrix}$ are full-rank, and $t < m$. Let $F(\underline{\beta}) = \sum_{i=1}^{t} |b_i - \underline{a}_i^T\underline{\beta}|^p$. By Theorem 2.2.2, F is convex. Also, $\frac{\partial}{\partial\underline{\beta}} F(\underline{\beta})$ exists and is continuous. Let $f(\underline{\beta}) = \sum_{i=1}^{n} |y_i - \underline{x}_i^T\underline{\beta}|^p$. By Theorem 2.2.2, f is strictly convex. Also, $\frac{\partial}{\partial\underline{\beta}} f(\underline{\beta})$ exists and is continuous. We assume that $D_c = \{\underline{\beta} \in \mathbb{R}^m | f(\underline{\beta}) \leq c\}$ is bounded, where $c \in \mathbb{R}$. Let $\underline{\beta}^o$ be the unique vector in $\mathbb{R}^m$ such that $f(\underline{\beta}^o) \leq f(\underline{\beta})$ for all $\underline{\beta} \in \mathbb{R}^m$. Note that $S^* = \{\underline{\beta}^* \in \mathbb{R}^m | A\underline{\beta}^* = \underline{b}\}$. Hence, $S^* \neq \emptyset$. $\underline{\beta}^o \notin S^*$ since otherwise the restrictions are redundant and we will have an unconstrained ℓ_p estimation problem. There is a unique $\underline{\beta}_\lambda \in \mathbb{R}^m$ for every $\lambda > 0$ by Lemma 2.3.2 and the fact that $F + \lambda f$ is strictly convex. Also, F is continuous implies B_ρ is closed and nonempty when $\rho > M$, then by Lemma 2.3.3, there exists a $\tilde{\underline{\beta}} \in B_\rho$ such that $f(\tilde{\underline{\beta}}) \leq f(\underline{\beta})$ for all $\underline{\beta} \in B_\rho$. Therefore,

we have all conditions needed for Theorem 5.6.1. Hence, by Theorem 5.6.1, there exists a unique $\underline{\beta}^{**} \in S^*$, i.e., $A\underline{\beta}^{**} = \underline{b}$, such that $f(\underline{\beta}^{**}) \leq f(\underline{\beta}^*)$ for all $\underline{\beta}^* \in S^*$, i.e., $A\underline{\beta}^* = \underline{b}$, and $\lim_{\lambda \to 0} \underline{\beta}_\lambda = \underline{\beta}^{**}$. Thus, we can find the constrained ℓ_p estimate by utilizing a sequence of unconstrained ℓ_p estimates.

If we are interested in finding the ℓ_p estimate having the least ℓ_p norm for the linear model in Section 1.2 when X is not full-rank and $p > 1$, we can proceed similarly. As indicated in Section 2.1, $\underline{\beta}^*$ is an ℓ_p estimate if and only if $U_1^T\underline{\beta}^* = \underline{a}_1^*$, where U_1^T and $\underline{a}_1^*$ are well-defined constants. Then, let $F(\underline{\beta}) = \sum_{i=1}^{t} |a_i - \underline{U}_i^T\underline{\beta}|^p$, where $\underline{a}_1^* = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_t \end{pmatrix}$ and $U_1^T = \begin{pmatrix} \underline{U}_1^T \\ \underline{U}_2^T \\ \vdots \\ \underline{U}_t^T \end{pmatrix}$ and let $f(\underline{\beta}) = \sum_{j=1}^{m} |\beta_j|^p$. Note that $D_c = \{\underline{\beta} \in \mathbb{R}^m \mid \|\beta\|_p \leq c^{\frac{1}{p}}\}$, hence, it is bounded for all $c \in \mathbb{R}$. Let $\underline{\beta}_\lambda \in \mathbb{R}^m$ such that $F(\underline{\beta}_\lambda) + \lambda f(\underline{\beta}_\lambda) \leq F(\underline{\beta}) + \lambda f(\underline{\beta})$ for all $\underline{\beta} \in \mathbb{R}^m$. By Theorem 5.6.1, $\lim_{\lambda \to 0} \underline{\beta}_\lambda = \underline{\beta}^{**}$, where $\underline{\beta}^{**}$ is the ℓ_p estimate having the least ℓ_p norm. Note that using this F function instead of the F function in Section 2.1 would significantly save the amount of computation for $\underline{\beta}_\lambda$.

In fact, there is a direct approach to finding the ℓ_p estimate having the least ℓ_p norm for the model we just considered. Let us

consider the linear model fitting problem $\underline{z} \cong W\underline{\tau}$. The ℓ_p estimation problem, $p > 1$, is to find an $\tilde{\underline{e}} \in \mathbb{R}^m$ such that $\underline{z} = W\tilde{\underline{\tau}} + \tilde{\underline{e}}$ and $\|\tilde{\underline{e}}\|_p^p \leq \|\underline{e}\|_p^p$ for all $\underline{e} \in \mathbb{R}^m$ such that $\underline{z} = W\underline{\tau} + \underline{e}$. That is, to find an $\tilde{\underline{e}} \in \mathbb{R}^m$ such that $W^{\perp T}\tilde{\underline{e}} = W^{\perp T}\underline{z}$ and $\|\tilde{\underline{e}}\|_p^p \leq \|\underline{e}\|_p^p$ for all $\underline{e} \in \mathbb{R}^m$ such that $W^{\perp T}\underline{e} = W^{\perp T}\underline{z}$, where $W^{\perp}$ is the orthogonal column space of W . Now, we want to find a $\tilde{\underline{\beta}} \in \mathbb{R}^m$ such that $U_1^T\tilde{\underline{\beta}} = \underline{a}_1^*$ and $\|\tilde{\beta}\|_p^p \leq \|\beta\|_p^p$ for all $\underline{\beta} \in \mathbb{R}^m$ such that $U_1^T\underline{\beta} = \underline{a}_1^*$. Since $\begin{pmatrix} U_1^T \\ U_2^T \end{pmatrix} (U_1 U_2) = I$, $U_1^T U_1 = I$ and $U_1 = U_2^{\perp}$. Therefore, $U_1^T\underline{\beta} = \underline{a}_1^*$ if and only if $U_2^{\perp T}\underline{\beta} = U_2^{\perp T}U_1\underline{a}_1^*$. Hence, $\underline{\beta}$ is the residual vector of the ℓ_p estimate for the linear model fitting $\underline{z} \cong W\underline{\tau}$, where $\underline{z} = U_1\underline{a}_1^*$ and $W = U_2$. Note that U_2 is full-rank, hence we should only get one such $\tilde{\underline{\beta}}$.

6. ℓ_p ESTIMATION IN THE CONSTRAINED LINEAR MODEL UNDER LINEAR INEQUALITY RESTRICTIONS

Armstrong and Frome (1976) studied computation of least squares estimates in the linear model when the parameter are restricted to be nonnegative. They proposed a branch-and-bound algorithm which requires solution of only a relatively small number of unrestricted least squares subproblems to obtain the solution of the restricted least squares problem. Khuri (1976) demonstrated a technique of transforming general restricted least squares problems to restricted least squares problems having nonnegative restrictions only. Waterman (1977) argued that the number of unrestricted least squares subproblems required to solve in the branch-and-bound algorithm proposed by Armstrong and Frome (1976) can be further reduced. Gentle and Kennedy (1979) studied the linear model with linear restrictions using various criteria for estimation, such as least squares, ℓ_1, M-estimation, etc. Also, they discussed the case in which the parameters are restricted to intervals and proposed a branch-and-bound method to solve the problem.

We will concentrate on the constrained ℓ_p estimation problem under linear inequality restrictions as discussed in Chapter 1. In case $p = 1$, it seems most appropriate to transform the problem to a linear programming problem with additional linear restrictions. Hence, we can solve the constrained ℓ_1 estimation problem essentially in the same

manner as we solve the unconstrained ℓ_1 estimation problem. In this chapter, we will discuss the constrained ℓ_p estimation problem for $p > 1$. The reparametrization of the constrained ℓ_p estimation problem will be given in Section 6.1. The details of a branch-and-bound method for the constrained ℓ_p estimation problem having nonnegative restrictions only will be provided in Section 6.2.

6.1 Reparametrization of the Problem

Khuri (1976) reparametrized general restricted least squares problems and formed restricted least squares problems having nonnegative restrictions only. The technique of reparameterization is also applicable to constrained ℓ_p estimation problems. We now use the technique to transform the constrained ℓ_p estimation problem discussed in Chapter 1 to a constrained ℓ_p estimation problem having nonnegative restrictions only.

Let us now consider the constrained ℓ_p estimation problem given in Chapter 1. Thus, we need to minimize $\|\underline{y} - X\underline{\beta}\|_p$ for all $\underline{\beta}$ such that $A\underline{\beta} \geq \underline{b}$, where A is a matrix of dimension $r \times m$ and $r \leq m$. We will assume a full-rank A here. Let A be decomposed as $A = (A_1 \quad A_2)$, where A_1 and A_2 are matrices of dimension $r \times r$ and $r \times (m-r)$, respectively. Without loss of generality, we assume that A_1 is nonsingular. Let X be decomposed accordingly as $X = (X_1 \quad X_2)$, where X_1 and X_2 are matrices of dimension $n \times r$ and $n \times (m-r)$, respectively. Let $\underline{\beta}$ be decomposed according as

$\underline{\beta} = \begin{bmatrix} \underline{\beta}^{(1)} \\ \underline{\beta}(2) \end{bmatrix}$, where $\underline{\beta}^{(1)}$ and $\underline{\beta}^{(2)}$ are vectors of dimension r and m-r , respectively. Now, let $\underline{\alpha}^{(1)} = A\underline{\beta} - \underline{b}$. Since $\underline{\alpha}^{(1)} = A_1\underline{\beta}^{(1)} + A_2\underline{\beta}^{(2)} - \underline{b}$, we have $\underline{\beta}^{(1)} = A_1^{-1}(\underline{\alpha}^{(1)} + \underline{b} - A_2\underline{\beta}^{(2)})$.

Hence,

$$\begin{aligned} \underline{y} - X\underline{\beta} &= \underline{y} - X_1\underline{\beta}^{(1)} - X_2\underline{\beta}^{(2)} \\ &= \underline{y} - X_1A_1^{-1}(\underline{\alpha}^{(1)} + \underline{b} - A_2\underline{\beta}^{(2)}) - X_2\underline{\beta}^{(2)} \\ &= \underline{y} - X_1A_1^{-1}\underline{b} - X_1A_1^{-1}\underline{\alpha}^{(1)} - (X_2 - X_1A_1^{-1}A_2)\underline{\beta}^{(2)} \\ &= \underline{z} - W\begin{bmatrix} \underline{\alpha}^{(1)} \\ \underline{\beta}^{(2)} \end{bmatrix} \\ &= \underline{z} - W\underline{\alpha} \end{aligned}$$

where $\underline{z} = \underline{y} - X_1A_1^{-1}\underline{b}$, $W = (X_1A_1^{-1}, X_2 - X_1A_1^{-1}A_2)$, $\underline{\alpha} = \begin{bmatrix} \underline{\alpha}^{(1)} \\ \underline{\alpha}(2) \end{bmatrix}$, and $\underline{\alpha}^{(2)} = \underline{\beta}^{(2)}$. Note that $\underline{z}$ is of dimension n and W is of dimension n x m . Hence, after the reparameterization of the constrained ℓ_p estimation problem, we form a constrained ℓ_p estimation problem having nonnegative restrictions only. In other words, we now minimize $\|\underline{z} - W\underline{\alpha}\|_p$ for all $\underline{\alpha}$ such that $\underline{\alpha}^{(1)} \geq \underline{0}$, where $\underline{\alpha} = \begin{bmatrix} \underline{\alpha}^{(1)} \\ \underline{\alpha}(2) \end{bmatrix}$. Once the optimal $\underline{\alpha}$ is obtained, the solution $\underline{\beta}$ for the original constrained ℓ_p estimation problem can be retrieved by

$$\underline{\beta} = \begin{bmatrix} \underline{\beta}^{(1)} \\ \underline{\beta}(2) \end{bmatrix} ,$$

$$\underline{\beta}^{(1)} = A_1^{-1}(\underline{\alpha}^{(1)} + \underline{b} - A_2\,\underline{\alpha}^{(2)})$$

and

$$\underline{\beta}^{(2)} = \underline{\alpha}^{(2)} \quad .$$

Note that, since

$$(X_1 \quad X_2)\begin{pmatrix} A_1^{-1} & -A_1^{-1}A_2 \\ 0 & I \end{pmatrix} = W$$

and

$$\begin{pmatrix} A_1 & A_2 \\ 0 & I \end{pmatrix}\begin{pmatrix} A_1^{-1} & -A_1^{-1}A_2 \\ 0 & I \end{pmatrix} = \begin{pmatrix} A_1^{-1} & -A_1^{-1}A_2 \\ 0 & I \end{pmatrix}\begin{pmatrix} A_1 & A_2 \\ 0 & I \end{pmatrix} = I$$

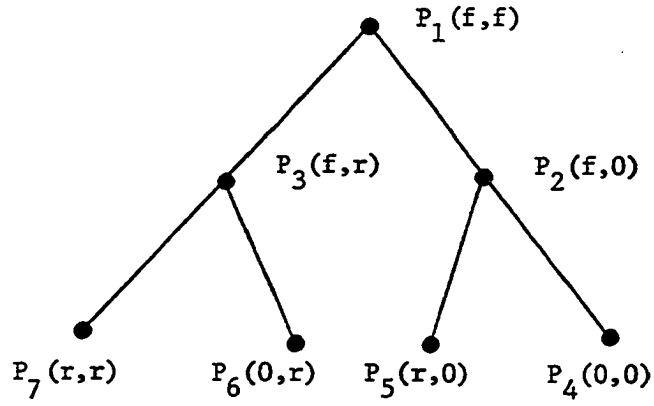
we have rank (W) = rank (X) . Thus, W is full-rank if and only if X is full-rank.

6.2 Branch-and-Bound Method

We now concentrate on solving the problem of minimizing $\|\underline{z} - W\underline{\alpha}\|_p$ for all $\underline{\alpha}$ such that $\underline{\alpha}^{(1)} \geq 0$, where $\underline{\alpha} = \begin{pmatrix} \underline{\alpha}^{(1)} \\ \underline{\alpha}^{(2)} \end{pmatrix}$, $\underline{\alpha}^{(1)}$ is of dimension r , and $\underline{\alpha}^{(2)}$ is of dimension m-r . Let $\underline{\alpha}^*$ be such that $\|\underline{z} - W\underline{\alpha}^*\|_p \leq \|\underline{z} - W\underline{\alpha}\|$ for all $\underline{\alpha} \in \mathbb{R}$. If $\underline{\alpha}^{(1)*} \geq 0$, then $\underline{\alpha}^*$ is the optimal solution of the problem. Otherwise, the optimal solution of the problem must lie on the boundary of the constraints $\underline{\alpha}^{(1)} \geq 0$. Suppose $\tilde{\underline{\alpha}}$ is the optimal solution of the problem such that $\tilde{\underline{\alpha}}^{(1)} = \begin{pmatrix} \tilde{\underline{\alpha}}_1^{(1)} \\ \tilde{\underline{\alpha}}_2^{(1)} \end{pmatrix}$, $\tilde{\underline{\alpha}}_1^{(1)} = \underline{0}$, and $\tilde{\underline{\alpha}}_2^{(1)} > \underline{0}$. Then, $\tilde{\underline{\alpha}}$ is also the solution

of the problem minimizing $\|\underline{z} - W\underline{\alpha}\|_p$ for all $\underline{\alpha}$ such that $\underline{\alpha}_1^{(1)} = \underline{0}$. The constraints $\underline{\alpha}_1^{(1)} \geq 0$ are called the active constraints of the original problem. Note that, setting $\underline{\alpha}_1^{(1)} = \underline{0}$ in the fitting equation $W\underline{\alpha} \cong \underline{z}$, we are really solving an unconstrained ℓ_p estimation subproblem. Also, if we assume a full-rank X, W is full-rank. Hence, the above unconstrained subproblem is a full-rank problem. As indicated in Section 2.3, we have a unique solution for the subproblem. There are 2^r unconstrained ℓ_p estimation subproblems under consideration. We will discuss a branch-and-bound method which requires to solve only a small portion of the complete set of subproblems.

A binary tree is constructed in the branch-and-bound method. Each node of the tree represents an unconstrained ℓ_p estimation subproblem. We use the notation $P_j(f, 0, \ldots)$ to denote the subproblem for node j in which the first parameter is free, the second parameter is set to zero, and so on. One daughter node is formed by setting a specific free parameter of the parent node to zero. The other daughter node is formed by placing "r" in the corresponding parameter position to indicate that the parameter is left unrestricted but required in the model. Note that, in this notation, "r" is used for the purpose of constructing the binary tree, we do not distinguish between "f" and "r" while we identify each node for the subproblem it represents. For example, we depict a full binary tree for the case $r = 2$ as follows.



Note that, $P_1(f,f)$, $P_3(f,r)$, and $P_7(r,r)$ represent the same unconstrained problem which we start with, and $P_2(f,0)$ and $P_5(r,0)$ represent the same unconstrained subproblem having 0 on the second parameter. Also, there are 4 distinct unconstrained subproblems in this example.

Let $\underline{\alpha}_j$ be the solution for the subproblem corresponding to node j in the binary tree. Let $R(P_j) = \|\underline{z} - W\underline{\alpha}_j\|_p$. We do not need to solve all 2^r unconstrained subproblems since some nodes in the binary tree may be fathomed (no descendant nodes formed). Note that, $R(P_k \geq R(P_j)$ if P_k is a descendant of P_j . Hence, we can fathom node j if (a) $\underline{\alpha}_j$ satisfies the constraints or (b) $R(P_j)$ is larger than R^* , the minimal upper bound available for the ℓ_p residual norm of the final solution. Such upper bound is always available, for example $R(P(0, 0, \ldots, 0))$.

Waterman (1977) discovered a property which provides powerful fathoming capabilities in a branch-and-bound method for restricted least squares problems having nonnegative restrictions only. The

property can be easily extended to our case. We now state the property and provide a proof which is similar to the one given by Waterman (1977).

Theorem 6.2.1. Let $\underline{\alpha}^*$ be such that $\|\underline{z} - W\underline{\alpha}^*\|_p \leq \|\underline{z} - W\underline{\alpha}\|_p$ for all $\underline{\alpha} \in \mathbb{R}^m$. Let $\tilde{\underline{\alpha}}$ be such that $\tilde{\underline{\alpha}}^{(1)} \geq \underline{0}$ and $\|\underline{z} - W\tilde{\underline{\alpha}}\|_p \leq \|\underline{z} - W\underline{\alpha}\|_p$ for all $\underline{\alpha}$ such that $\underline{\alpha}^{(1)} \geq \underline{0}$. If k $(1 \leq k \leq r)$ of the components of $\underline{\alpha}^{(1)*}$ are negative, then at least one of the corresponding k components in $\underline{\alpha}^{(1)}$ is zero.

Proof: Without loss of generality, we assume the first k components of $\underline{\alpha}^{(1)*}$ are negative. In other words, $\alpha_i^{(1)*} < 0$ for $1 \leq i \leq k$ and $\alpha_i^{(1)*} \geq 0$ for $k < i \leq r$. Suppose the theorem is not true, then we have $\tilde{\alpha}_i^{(1)} > 0$ for $1 \leq i \leq k$ and $\tilde{\alpha}_i^{(1)} \geq 0$ for $k < i \leq r$. Hence, there exists $0 < \lambda_i < 1$ such that $\lambda_i \alpha_i^{(1)*} + (1 - \lambda_i)\tilde{\alpha}_i^{(1)} > 0$ for $1 \leq i \leq k$. Let $\lambda = \min(\lambda_1, \lambda_2, \ldots, \lambda_k)$. Let $\tau_i = \lambda\alpha_i^* + (1 - \lambda)\tilde{\alpha}_i$ for $1 \leq i \leq m$. Then, $\tau_i \geq 0$ for $1 \leq i \leq r$. Now,

$$\begin{aligned}
\|\underline{z} - W\underline{\tau}\|_p &= \|(\lambda\underline{z} - \lambda W\underline{\alpha}^*) + ((1-\lambda)\underline{z} - (1-\lambda)W\tilde{\underline{\alpha}})\|_p \\
&\leq \|\lambda(\underline{z} - W\underline{\alpha}^*))\|_p + \|(1-\lambda)(\underline{z} - W\tilde{\underline{\alpha}})\|_p \\
&= \lambda\|\underline{z} - W\underline{\alpha}^*\|_p + (1-\lambda)\|\underline{z} - W\tilde{\underline{\alpha}}\|_p \\
&\leq \lambda\|\underline{z} - W\tilde{\underline{\alpha}}\|_p + (1-\lambda)\|\underline{z} - W\tilde{\underline{\alpha}}\|_p \\
&= \|\underline{z} - W\tilde{\underline{\alpha}}\|_p \quad .
\end{aligned}$$

Note that, we have used the property of triangular inequality and the fact that $\underline{\alpha}^*$ is the solution of the unconstrained problem. Now, if

$\|\underline{z} - W\underline{\tau}\|_p < \|\underline{z} - W\underline{\tilde{\alpha}}\|_p$ then, since $\tau_i \geq 0$ for $1 \leq i \leq r$, this contradicts the fact that $\underline{\tilde{\alpha}}$ is the solution of the constrained problem. If $\|\underline{z} - W\underline{\tau}\|_p = \|\underline{z} - W\underline{\tilde{\alpha}}\|_p$, then, since $\underline{\tau} \neq \underline{\tilde{\alpha}}$, this contradicts the uniqueness of solution for the constrained problem if we assume a full-rank X as indicated earlier. Therefore, it is wrong to suppose the theorem is not true. Hence, we have proved the theorem. □

The immediate result of the theorem is the following fathoming capabilities. Suppose, at node j, $\underline{\alpha}_1^{(1)}$ is the subvector of the solution $\underline{\alpha}_j$ consisting of the parameters which violate the constraints, i.e., $\underline{\alpha}_1^{(1)} < 0$. If there is a solution (to the original constrained problem) among the descendants of node j, then the solution has zero on at least one of the parameters corresponding to parameters of $\underline{\alpha}_1^{(1)}$. Further, fathoming capabilities are given by the following theorem, similar to the one given by Armstrong and Frome (1976).

Theorem 6.2.2. Suppose, at node j, the component $\alpha_i^{(1)}$ in the solution $\underline{\alpha}_j$ violates the constraint, i.e., $\alpha_i^{(1)} < 0$. Let P_k be the subproblem formed by setting $\alpha_i^{(1)} = 0$. Let $\underline{\alpha}_k$ be the solution of P_k. If $\underline{\alpha}_k$ satisfies the constraints or $R(P_k) \geq R^*$, both node k and its sister node can be fathomed.

Proof: We will first prove that $\underline{\alpha}_k$ is actually the solution of the problem formed by setting $\alpha_i^{(1)} \geq 0$ instead of $\alpha_i^{(1)} = 0$. Let us concentrate on the subproblem P_j. As indicated earlier, P_j is an unconstrained ℓ_p estimation problem having some parameters fixed to be

zeros. $\underline{\alpha}_j$ is the solution of the unconstrained problem such that the component $\alpha_i^{(1)}$ of $\underline{\alpha}_j$ is negative. Let $\tilde{\underline{\alpha}}$ be the solution of the constrained problem formed by adding the constraint $\alpha_i^{(1)} \geq 0$ to the unconstrained problem. By Theorem 6.2.1, we have $\tilde{\alpha}_i^{(1)} = 0$. Since $\underline{\alpha}_k$ is the solution of the subproblem P_k formed by setting $\alpha_i^{(1)} = 0$, we have $\tilde{\underline{\alpha}} = \underline{\alpha}_k$. In other words, $\underline{\alpha}_k$ is the solution of the problem formed by setting $\alpha_i^{(1)} \geq 0$.

Now, suppose that in some descendant of the sister node of node k, a solution $\underline{\alpha}_q$ is obtained such that all constraints are satisfied. Since all the restrictions of the parent problem are still present, $\underline{\alpha}_q$ satisfies all these restrictions and $\alpha_i^{(1)} \geq 0$. Hence, $R(\underline{\alpha}_q) \geq R(\underline{\alpha}_k)$. Therefore, if $\underline{\alpha}_k$ satisfies the constraints or $R(P_k) \geq R^*$, both node k and its sister node can be fathomed. □

We would also like to check the Kuhn-Tucker conditions of the original constrained problem at the feasible solution of each subproblem. We can claim the feasible solution is the optimal solution of the constrained problem if the Kuhn-Tucker conditions are satisfied. We now derive these conditions as follows. Let $F(\underline{\alpha}) = \sum_{i=1}^{n} |z_i - \underline{w}_i^T \underline{\alpha}|^p$, where $\underline{w}_i^T$ is the i^{th} row of W. The problem of interest is to minimize $F(\underline{\alpha})$ such that $\underline{\alpha}^{(1)} \geq 0$, where $\underline{\alpha} = \begin{pmatrix} \underline{\alpha}^{(1)} \\ \underline{\alpha}^{(2)} \end{pmatrix}$. Let

$\phi(\underline{\alpha}, \underline{\lambda}) = F(\underline{\alpha}) - \underline{\lambda}^T \underline{\alpha}^{(1)}$. The Kuhn-Tucker conditions of the problem are given as follows:

(a) $\frac{\partial}{\partial \lambda} \phi(\underline{\alpha}_0, \underline{\lambda}_0)^T \underline{\lambda}_0 = 0$,

(b) $\frac{\partial}{\partial \lambda} \phi(\underline{\alpha}_0, \underline{\lambda}_0) \leq \underline{0}$,

(c) $\underline{\lambda}_0 \geq 0$,

(d) $\frac{\partial}{\partial \underline{\alpha}} \phi(\underline{\alpha}_0, \underline{\lambda}_0) = \underline{0}$

These conditions can be written as follows:

(a) $\underline{\alpha}_0^{(1)^T} \underline{\lambda}_0 = 0$,

(b) $\underline{\alpha}_0^{(1)} \geq 0$,

(c) $\underline{\lambda}_0 \geq 0$,

(d) $\frac{\partial}{\partial \underline{\alpha}} F(\underline{\alpha}_0) = \begin{pmatrix} \underline{\lambda}_0 \\ \underline{0} \end{pmatrix}$.

We abbreviate these conditions as follows:

$$\text{for } 1 \leq i \leq r \text{ , if } \alpha_i^{(1)} > 0 \text{ , then } \left(\frac{\partial}{\partial \underline{\alpha}} F(\underline{\alpha})\right)_i = 0$$

$$\text{if } \alpha_i^{(1)} = 0 \text{ , then } \left(\frac{\partial}{\partial \underline{\alpha}} F(\underline{\alpha})\right)_i \geq 0 \text{ ;}$$

$$\text{for } r < i \leq m \text{ , } \left(\frac{\partial}{\partial \underline{\alpha}} F(\underline{\alpha})\right)_i = 0 \text{ .}$$

We summarize the branch-and-bound method as follows. Form the root of a binary tree having free parameter for every parameter in the constraints. Check the negative components of the solution and create daughter nodes by setting a corresponding free parameter to zero. Obtain an R^* (an upper bound on ℓ_p residual norm). Fathom node j if (a) $\underline{\alpha}_j$ satisfies the constraints, (b) $R(P_j) \geq R^*$, or (c) sister

node is fathomed (an application of Theorem 6.2.2). For each solution $\underline{\alpha}_k$ satisfying the constraints, update R^* if $R(P_k) \leq$ Previous R^*, also check the Kuhn-Tucker conditions. If the Kuhn-Tucker conditions are satisfied, stop; else continue the development of the binary tree.

7. APPENDIX

```
C  PROGRAM TO COMPUTE LP ESTIMATE IN THE LINEAR MODEL
C  Y = X * B + E, P > 1 AND P ¬= 2.
C
C  SUBROUTINE GSWP AND TRAN ARE REQUIRED.
C
      REAL*8 Y(50),X(50,10),B(10),E(50),S1(10,10),S2(10,50),S3(50,50),
     1S4(10),LSB(10),LSR(50),WO(50),T(50),U(50),W(50),Z(50),SV(11,11),
     2R(50),RBAR(50),WBAR(50),G(10),F,XI,XIPR,EPS,EPSPR,ALPHA,P,PZ,PW
C
      NN=50
      MM=10
C                                   READ P,XI,XIPR,EPS,EPSPR
C
      READ(5,8,END=999)P,XI,XIPR,EPS,EPSPR
    8 FORMAT(F6.2,4D10.2)
C
      WRITE(6,9)P,XI,XIPR,EPS,EPSPR
    9 FORMAT('1','P = ',F6.2,/,' XI = ',D10.2,/,' XIPR = ',D10.2,/,
     1' EPS = ',D10.2,/,' EPSPR = ',D10.2)
C
C                                   COMPUTE PZ = P - 1
      PZ=P-1.D0
C                                   COMPUTE PW = 1 / (P - 1)
      PW=1.D0/PZ
C                                  Y IS OF DIMENSION N TIMES 1,
C                                  X IS OF DIMENSION N TIMES M.
      N=12
      M=5
C                                  READ Y AND X
C
      READ(5,10,END=999)((Y(I),X(I,J),J=1,M),I=1,N)
   10 FORMAT(6F10.4)
C
      WRITE(6,20)
   20 FORMAT('-','THE 1ST COL IS FOR Y AND THE REST IS FOR X.',/)
      WRITE(6,30)((Y(I),X(I,J),J=1,M),I=1,N)
   30 FORMAT('0',G18.8,5X,5G18.8)
C                                   COMPUTE S1 = INV(X' * X)
      DO 110 I=1,M
      DO 110 J=I,M
        F=0.D0
        DO 100 K=1,N
  100     F=F+X(K,I)*X(K,J)
        S1(I,J)=F
  110   S1(J,I)=F
      I=1
      RANK=0
      CALL GSWP(S1,I,M,MM,RANK)
      TR=RANK
```

```
C                                       COMPUTE S2 = INV(X' * X) * X'
      DO 130 I=1,M
      DO 130 J=1,N
        F=0.D0
        DO 120 K=1,M
 120      F=F+S1(I,K)*X(J,K)
 130    S2(I,J)=F
C                                       COMPUTE S3 = I - X * INV(X' * X) * X'
      DO 150 I=1,N
      DO 150 J=I,N
        F=0.D0
        DO 140 K=1,M
 140      F=F+X(I,K)*S2(K,J)
        IF (I.EQ.J) S3(I,J)=1.D0-F
        IF (I.NE.J) S3(I,J)=-F
 150    S3(J,I)=S3(I,J)
C                                       COMPUTE S4 = X' * Y
      DO 170 I=1,M
        F=0.D0
        DO 160 K=1,N
 160      F=F+X(K,I)*Y(K)
 170    S4(I)=F
C                                       START WITH LEAST SQUARES SOLUTION
      DO 210 I=1,M
        F=0.D0
        DO 200 K=1,N
 200      F=F+S2(I,K)*Y(K)
 210    LSB(I)=F
C
      DO 230 I=1,N
        F=0.D0
        DO 220 K=1,M
 220      F=F+X(I,K)*LSB(K)
 230    LSR(I)=Y(I)-F
C
      CALL TRAN(LSR,W0,N,NN,PZ)
 250  CONTINUE
C                                       ITERATIVE PROCEDURE STARTS HERE
C
C                                       COMPUTE U
      DO 310 I=1,M
        F=0.D0
        DO 300 K=1,N
 300      F=F+S2(I,K)*W0(K)
 310    T(I)=F
      DO 330 I=1,N
        F=0.D0
        DO 320 K=1,M
 320      F=F+X(I,K)*T(K)
 330    U(I)=-F
```

```
      IF (P.LE.2.D0) GO TO 350
      DO 340 I=1,N
  340 IF (DABS(W0(I)).LE.XI) U(I)=0.D0
  350 CONTINUE
C                                     COMPUTE W
      DO 360 I=1,N
  360 W(I)=W0(I)+EPSPR*U(I)
      CALL TRAN(W,Z,N,NN,PW)
C                                     OBTAIN INV(V)
      M1=M+1
      DO 370 I=1,M
      DO 370 J=1,M
  370 SV(I,J)=S1(I,J)
      DO 390 I=1,M
      F=0.D0
        DO 380 K=1,N
  380     F=F+S2(I,K)*Z(K)
        SV(I,M1)=F
  390   SV(M1,I)=-F
      DO 410 I=1,N
        F=0.D0
        DO 400 K=1,N
  400     F=F+S3(I,K)*Z(K)
  410   T(I)=F
      F=0.D0
      DO 420 I=1,N
  420   F=F+Z(I)*T(I)
      SV(M1,M1)=F
      RANK=TR
      CALL GSWP(SV,M1,M1,MM,RANK)
C                                     COMPUTE B AND ALPHA
      F=0.D0
      DO 430 I=1,N
  430   F=F+Z(I)*Y(I)
      T(1)=F
      DO 450 I=1,M
        F=0.D0
        DO 440 K=1,M
  440     F=F+SV(I,K)*S4(K)
  450   B(I)=F+SV(I,M1)*T(1)
      F=0.D0
      DO 460 I=1,M
  460 F=F+SV(M1,I)*S4(I)
      ALPHA=F+SV(M1,M1)*T(1)
C                                     COMPUTE E
      DO 480 I=1,N
         F=0.D0
        DO 470 K=1,M
  470     F=F+X(I,K)*B(K)
  480   E(I)=Y(I)-F-ALPHA*Z(I)
```

```
      IF (P.GE.2.D0) GO TO 500
      DO 490 I=1,N
 490    IF (DABS(ALPHA*Z(I)).LE.XI) E(I)=0.D0
 500  CONTINUE
C                                     COMPUTE R
      DO 510 I=1,N
 510    R(I)=ALPHA*Z(I)+EPS*E(I)
C                                     COMPUTE W0
      CALL TRAN(R,W0,N,NN,PZ)
C                                     COMPUTE RBAR
      DO 520 I=1,N
 520    RBAR(I)=R(I)+(1.D0-EPS)*E(I)
C                                     COMPUTE WBAR
      CALL TRAN(RBAR,WBAR,N,NN,PZ)
C                                     COMPUTE G, THE GRADIENT
      DO 540 I=1,M
        F=0.D0
        DO 530 K=1,N
 530      F=F+X(K,I)*WBAR(K)
 540    G(I)=-P*F
C                                     WRITE B AND STOP IF G IS CLOSE TO 0
      F=0.D0
      DO 550 I=1,M
 550    F=F+G(I)**2
      IF (DSQRT(F).GT.XIPR) GO TO 250
      WRITE(6,560)(B(I),I=1,M)
 560  FORMAT('1','THE LP ESTIMATE IS :',/,6G18.8)
 999  STOP
      END

      SUBROUTINE GSWP(A,M0,M,MM,RANK)
C
C  ROUTINE TO USE GENERALIZED SWEEPS TO FIND
C  GENERALIZED INVERSE OF GIVEN SYMMETRIC
C  MATRIX. ALSO FINDS RANK OF MATRIX.
C  IT WILL START AT ROW M0 AND END AT ROW M.
C
      REAL*8 A(MM,MM),C,D
      DO 10 K=M0,M
         D=A(K,K)
         IF(DABS(D).GE..1D-12) GO TO 30
         DO 20 J=1,M
            A(K,J)=0.D0
 20         A(J,K)=0.D0
         GO TO 10
 30      DO 40 J=1,M
 40         A(K,J)=A(K,J)/D
         RANK=RANK+1
         DO 50 I=1,M
            IF(I.EQ.K) GO TO 50
            C=A(I,K)
```

```
            DO 45 J=1,M
 45            A(I,J)=A(I,J)-C*A(K,J)
            A(I,K)=-C/D
 50         CONTINUE
         A(K,K)=1.D0/D
 10      CONTINUE
      WRITE(6,60) RANK
 60   FORMAT(' RANK=',F8.1)
      RETURN
      END


      SUBROUTINE TRAN(U,V,N,NN,PP)
C
C  ROUTINE TO GENERATE Z FROM W, OR W0 FROM R
C
      REAL*8 U(NN),V(NN),PP
      DO 10 I=1,N
        IS=1
        IF(U(I).LT.0.D0) IS=-1
 10     V(I)=(DABS(U(I))**PP)*IS
      RETURN
      END
```

8. BIBLIOGRAPHY

Abdelmalek, N. N. 1971. Linear ℓ_1 approximation for a discrete point set and ℓ_1 solution of overdetermined linear equations. JACM 18: 41-47.

Armstrong, R. D., and E. L. Frome. 1976. A branch-and-bound solution of a restricted least squares problem. Technometrics 18: 447-450.

Banks, S. C., and H. L. Taylor. 1980. A modification to the discrete ℓ_1 linear approximation algorithm of Barrodale and Roberts. SIAM J. Sci. Stat. Comput. 1: 187-190.

Barrodale, I., and F. D. K. Roberts. 1970. Applications of mathematical programming to ℓ_1 approximation. Pp. 447-464 in J. B. Rosen, ed. Nonlinear programming. Academic Press, New York, N.Y.

Barrodale, I., and F. D. K. Roberts. 1973. An improved algorithm for discrete ℓ_1 linear approximation. SIAM J. Numer. Anal. 10: 839-848.

Bloomfield, P., and W. Steiger. 1980. Least absolute deviations curve-fitting. SIAM J. Sci. Stat. Comput. 1: 390-301.

Ekblom, H. 1973. Calculation of linear best ℓ_p-approximation. BIT 13: 292-300.

Fletcher, R., and M. D. J. Powell. 1963. A rapidly convergent descent method for minimization. Computer J. 6: 163-168.

Forsythe, A. B. 1972. Robust estimation of straight line regression coefficients by minimizing p-th power deviations. Technometrics 14: 159-166.

Gentle, J. E., and W. J. Kennedy. 1979. Algorithms for linear regression with linear restrictions. Pp. 339-343 in J. F. Gentleman, ed. Proceedings of the Computer Science and Statistics: 12th Annual Symposium on the Interface, Univ. of Waterloo, Waterloo, Ontario, Canada.

Harter, H. L. 1977. Nonuniqueness of least absolute values regression. Comm. in Stat. A6(9): 829-838.

Harvey, A. C. 1978. On the unbiasedness of robust regression estimators. Comm. in Stat. A7(8): 779-783.

Hoerl, A. E., and R. W. Kennard. 1970. Ridge regression. Biased estimation for nonorthogonal problem. Technometrics 12: 55-67.

Huber, P. J. 1964. Robust estimation of a location parameter. Ann. Math. Stat. 35: 73-101.

Kennedy, W. J., and J. E. Gentle. 1978. Comparisons of algorithms for minimum ℓ_p norm linear regression. Pp. 373-378 in D. Hogben, ed. Proceedings of Computer Science and Statistics: 10th Annual Symposium on the Interface, U. S. Government Printing Office, Washington, D.C.

Kennedy, W. J., and J. E. Gentle. 1980. Statistical computing. Marcel Dekker, Inc., New York, N.Y. 591 pp.

Khuri, A. I. 1976. A constrained least-squares problem. Comm. in Stat. B5: 82-84.

Marquardt, D. W. 1970. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. Technometrics 12: 591-612.

Merle, G., and H. Spath. 1974. Computational experiences with discrete ℓ_p-approximation. Computing 12: 315-321.

Money, A. H., Affleck-Graves, J. F., Hart, M. L., and G. D. I. Barr. 1982. The linear regression model: ℓ_p norm estimation and the choice of p. Comm. in Stat. 11(1): 89-109.

Rice, J. R., and J. S. White. 1964. Norms for smoothing and estimation. SIAM Rev. 6: 243-256.

Sielken, R. L. Jr., and H. O. Hartley. 1973. Two linear programming algorithms for unbiased estimation of linear models. J. Am. Stat. Assoc. 68: 639-641.

Sposito, V. A. 1975. Linear and nonlinear programming. The Iowa State University Press, Ames, Iowa. 269 pp.

Sposito, V. A. 1982. On unbiased ℓ_p regression estimators. Unpublished technical report. Dept. of Stat., Iowa State University, Ames, Iowa.

Usow, K. H. 1967. On ℓ_1 approximation II: computation for discrete functions and discretization effects. SIAM J. Numer. Anal. 4: 233-244.

Waterman, M. S. 1977. Least squares with nonnegative regression coefficients. J. Stat. Comp. Sim. 6: 67-70.

9. ACKNOWLEDGEMENTS

I wish to acknowledge and thank Professor William J. Kennedy, under whose supervision this thesis was done, for his encouragement, help, and advice. Also, I thank Professor Vince A. Sposito for his invaluable suggestions to my thesis.

My late wife Nan-Sie had provided exceptional love and support which are keys to my progress. I am greatly indebted to her. I would also like to thank my parents for their constant support all these years and my late mother-in-law Ai-Lin for her kindness and love.

Mrs. Darlene Wicks has done an excellent job of typing. Special thanks are due for her remarkable speed. I also like to express my appreciation to the faculty and staff at Iowa State University for their general assistance.